

Multivariate approximation in total variation

A. D. Barbour¹, M. J. Luczak² & A. Xia³
 Universität Zürich, Queen Mary University of London
 & University of Melbourne

December 24, 2015

Abstract

The paper lays the framework for the discrete normal approximation in total variation of random vectors in \mathbb{Z}^d , using Stein's method. We derive an appropriate Stein equation, together with bounds on its solutions and their differences, and use them to formulate a general discrete normal approximation theorem. We illustrate the use of the method in three settings: sums of independent integer valued random vectors, equilibrium distributions of Markov population processes, and random vectors exhibiting an exchangeable pair. We conclude with an application to random colourings of regular graphs.

Keywords: Markov population process; multivariate approximation; total variation distance; infinitesimal generator; Stein's method

AMS subject classification: Primary 62E17; Secondary 62E20, 60J27, 60C05

Running head: Multivariate approximation

¹Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich; e-mail a.d.barbour@math.uzh.ch. Work begun while ADB was Saw Swee Hock Professor of Statistics at the National University of Singapore, carried out in part at the University of Melbourne and at Monash University, and supported in part by Australian Research Council Grants Nos DP120102728, DP120102398, DP150101459 and DP150103588

²School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK; e-mail malwinajluczak@gmail.com. Work carried out in part at the University of Melbourne, and supported by an EPSRC Leadership Fellowship, grant reference EP/J004022/2, and in part by Australian Research Council Grants Nos DP120102398 and DP150101459.

³School of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia; e-mail a.xia@ms.unimelb.edu.au. Work supported in part by Australian Research Council Grants Nos DP120102398 and DP150101459.

1 Introduction

The Stein–Chen method (Chen, 1975) enables the distribution of a sum W of indicator random variables to be approximated by a Poisson distribution in a wide variety of circumstances. In addition, it provides an estimate of the accuracy of the approximation, expressed in terms of the *total variation distance*. Such an approximation is very valuable, since it allows the approximation of the probability $\mathbb{P}[W \in A]$ of an arbitrary subset A of \mathbb{Z}_+ by a Poisson probability, and not just of sets A with ‘nice’ properties. By contrast, the distance classically used for quantifying normal approximation is the Kolmogorov distance, as in the Berry–Esseen theorem, and this measures the largest difference between the probabilities of half lines. Of course, this can easily be extended to (the unions of small numbers of) intervals, but gives no information at all about, for instance, the probability that W is even.

The Poisson family of distributions is, however, too restrictive to be used as widely as the normal distribution for approximation, because mean and variance have to be equal. Starting from the seminal paper of Presman (1983), more general approximations in total variation have been derived, using more flexible families. In particular, for the translated Poisson family, the Stein–Chen method can be adapted in a natural way (Röllin 2005, 2007), allowing for the possibility of treating sums of dependent indicator random variables. What is more, the order of the error in total variation approximation obtained in this way, using the translated Poisson family (Barbour & Xia, 1999) or the discretized normal family (Fang, 2014), need be no worse than that of the error in the normal approximation, measured using Kolmogorov distance. This represents a substantial gain in the scope of the approximation, at relatively small cost.

In this paper, we derive analogous results in higher dimensions, an undertaking of considerably greater difficulty. For the classical multivariate central limit theorem, the accuracy of approximation is usually evaluated by comparing the worst approximation of the probability of a *convex* set, a natural d -dimensional analogue of the interval. In particular, it is important in the proofs of such results to be able to bound the probabilities of thin shells around the boundaries of the sets A under consideration, for which reason ‘nice’ sets are important. In our theorems, the probabilities of all sets, no matter how complicated their boundaries, are approximated. It is thus to be expected that a substantial loss of accuracy might result. It turns out that this is not the case; for a sum of n independent and identically distributed random vectors, for instance, the error rate that we establish is of order $O(n^{-1/2} \log n)$, worse only because of the logarithmic factor than

the ideal $O(n^{-1/2})$ rate of the Berry–Esseen theorem. What is more, the dependence of the error on the dimension d is still polynomial, with growth in d of at worst $O(d^4)$; for sums of independent and identically distributed random vectors and the convex set metric, the growth of the error bound with d has been shown to be of order $O(d^{7/4})$ in Bentkus (2003). However, there are other difficulties involved in total variation approximation on the d -dimensional lattice, which do not appear in the classical central limit setting. For instance, a random vector W with finite covariance matrix can always be normalized to have the identity as covariance matrix. Such a normalization, however, violates the lattice structure, so that the detail of the covariance matrix enters into the problem, and indeed into the error estimates, in a much more complicated way.

The first step in deriving discrete multivariate approximation is to choose a suitable family of reference distributions. It is easy enough to define (translated) multivariate Poisson families, if the correlations between their components are all non-negative. However, in a general multivariate setting, such an assumption is unnatural. A discretization of the family of multivariate normal distributions offers a much more flexible option. However, for this family, there is no pre-existing Stein operator or Stein equation to replace the Stein equations for the Poisson and normal distributions that are used heavily in the 1-dimensional case. We thus need to find a suitable multivariate operator.

For the Poisson distribution $\text{Po}(\lambda)$, there is a Markov population process, the immigration–death process with constant immigration rate λ and unit *per capita* death rate, whose equilibrium distribution is exactly $\text{Po}(\lambda)$, and whose generator can be used as the corresponding Stein operator (Barbour, 1988). Proceeding by analogy, we consider the equilibrium distributions of more general Markov population processes as possible reference distributions. As in the Poisson case, their generators automatically yield corresponding Stein equations (Barbour, Holst & Janson, Section 10.1). In addition, they come with a probabilistic representation of the solutions to the Stein equation that makes it possible to estimate the quantities needed in exploiting the method. Unfortunately, we usually have no readily available exact representation of the equilibrium distribution of a Markov population process, making their use as reference distributions inconvenient. However, when the spread of the equilibrium distribution of the Markov population process is large, we can show, under a weak irreducibility condition, that it is close in total variation to a discrete multivariate normal distribution. Since the error in this discrete normal approximation is of comparable order to that obtained when approximating the distribution of a sum of independent and identically distributed random vectors by the equilibrium distribution of a Markov population pro-

cess, the discrete multivariate normal can be used in place of the equilibrium distribution of the Markov population process, without altering the order of the error bound.

When using Stein's method to approximate the distribution of a random d -vector W , a key step is to show that, for a large class of functions $h: \mathbb{Z}^d \rightarrow \mathbb{R}$, the expectation $\mathbb{E}\{\mathcal{A}h(W)\}$ of the Stein operator \mathcal{A} applied to h is small. In our theorems, we use the functions h that are determined using the Stein equation for the equilibrium distribution of a Markov population process, in which the Stein operator \mathcal{A} is the generator of the Markov population process. However, we replace the Stein operator \mathcal{A} corresponding to the Markov population process with a simplified operator

$$\mathcal{A}_n h(w) := \frac{n}{2} \text{Tr}(\sigma^2 \Delta^2 h(w)) + \Delta h^T(w) A(w - nc), \quad w \in \mathbb{Z}^d, \quad (1.1)$$

where $c \in \mathbb{R}^d$, A is a matrix whose eigenvalues all have negative real parts, σ^2 is a positive definite symmetric matrix, and

$$\Delta_j h(w) := h(w + e^{(j)}) - h(w); \quad \Delta_{jk}^2 h(w) := \Delta_j(\Delta_k h)(w), \quad 1 \leq j, k \leq d.$$

The two operators are shown to be close to one another in Theorem 4.6. The operator \mathcal{A}_n is a discrete analogue of the d -dimensional Ornstein–Uhlenbeck operator

$$\mathcal{A}_\infty h(w) := \frac{n}{2} \text{Tr}(\sigma^2 D^2 h(w)) + Dh^T(w) A(w - nc), \quad w \in \mathbb{R}^d, \quad (1.2)$$

acting on twice differentiable functions $h: \mathbb{R}^d \rightarrow \mathbb{R}$. Note that we take all vectors, including $\Delta h(w)$ and $Dh(w)$, to be *column* vectors. The Ornstein–Uhlenbeck process corresponding to (1.2) has as equilibrium distribution the d -variate normal distribution $\mathcal{N}_d(nc, n\Sigma)$ with mean nc and covariance matrix $n\Sigma$, where Σ is the positive definite symmetric solution of the continuous Lyapounov equation

$$A\Sigma + \Sigma A^T + \sigma^2 = 0; \quad (1.3)$$

see, for example, Khalil (2002, Theorem 4.6, p.136).

Broadly speaking, we show that, if $\mathbb{E}\{\mathcal{A}_n h(W)\}$ is suitably small for enough functions h , then the distribution of W is close to a discrete normal distribution $\mathcal{DN}_d(nc, n\Sigma)$, obtained from $\mathcal{N}_d(nc, n\Sigma)$ by assigning the probability of the d -box

$$[i_1 - 1/2, i_1 + 1/2) \times \cdots \times [i_d - 1/2, i_d + 1/2)$$

to the integer (i_1, \dots, i_d) , for each $(i_1, \dots, i_d) \in \mathbb{Z}^d$. A precise statement of the approximation theorem is to be found in Theorem 1.1. There are many

choices of A and σ^2 that correspond to a given Σ , and the particular choice of A and σ^2 for use in (1.1) is typically dictated by the specific context. We are able to show that any such choice of A and σ^2 can be associated with a Markov population process X_n , whose generator, when simplified, reduces to that in (1.1).

The paper runs as follows. First, we begin with a sequence $(X_n, n \geq 1)$ of Markov population processes on \mathbb{Z}^d . X_n has transition rates of the following form:

$$X \rightarrow X + J \quad \text{at rate} \quad ng^J(n^{-1}X), \quad X \in \mathbb{Z}^d, J \in \mathcal{J}, \quad (1.4)$$

where \mathcal{J} is a finite subset of \mathbb{Z}^d , and the functions g^J are twice continuously differentiable on \mathbb{R}^d . We suppose that the equations

$$\frac{d\xi}{dt} = F(\xi) := \sum_{J \in \mathcal{J}} Jg^J(\xi) \quad (1.5)$$

have an equilibrium point c , so that $F(c) = 0$, and that the matrix

$$A := DF(c) \quad (1.6)$$

has eigenvalues whose real parts are all negative, making c a strongly stable equilibrium of (1.5). This matrix A will be the one that appears in the Stein operator (1.1). The matrix σ^2 that appears there is also expressed in terms of the transition rates of X_n :

$$\sigma^2 := \sigma^2(c), \quad \text{where} \quad \sigma^2(x) := \sum_{J \in \mathcal{J}} JJ^T g^J(x). \quad (1.7)$$

We now modify X_n to keep it reasonably close to nc . For technical reasons, we measure closeness in \mathbb{R}^d using the norm $|\cdot|_\Sigma$ defined by

$$|Y|_\Sigma^2 := Y^T \Sigma^{-1} Y, \quad (1.8)$$

where Σ is as defined above; we let $B_{\delta, \Sigma}(c) := \{\xi \in \mathbb{R}^d : |\xi - c|_\Sigma \leq \delta\}$. Defining

$$\mathcal{X}_n^\delta(J) := \{X \in \mathbb{Z}^d : \{X, X + J\} \subset B_{n\delta, \Sigma}(nc)\}, \quad (1.9)$$

we replace X_n with the process X_n^δ having transition rates

$$X \rightarrow X + J \quad \text{at rate} \quad ng_\delta^J(n^{-1}X) := \begin{cases} ng^J(n^{-1}X), & \text{if } X \in \mathcal{X}_n^\delta(J); \\ 0, & \text{otherwise,} \end{cases} \quad (1.10)$$

for $X \in \mathbb{Z}^d$ and $J \in \mathcal{J}$, with δ to be chosen suitably small and positive; broadly speaking, we choose δ so that c is a strongly attractive equilibrium of the equations (1.5) throughout $B_{\delta, \Sigma}(c)$. Then, if

$$X_n^\delta(0) \in \tilde{B}_{n, \delta}(c) := \mathbb{Z}^d \cap B_{n\delta, \Sigma}(nc), \quad (1.11)$$

it follows that X_n^δ is a Markov process on the finite state space $\tilde{B}_{n,\delta}(c)$, and so has an equilibrium distribution; furthermore, if all states in $\tilde{B}_{n,\delta}(c)$ communicate, this equilibrium distribution Π_n^δ is unique. Assumptions G3 and G4 below guarantee that this is the case: see Lemma 2.1.

Now, if $X_n^\delta \sim \Pi_n^\delta$, it follows by Dynkin's formula and because each set $\mathcal{X}_n^\delta(J)$ is bounded that $\mathbb{E}\{\mathcal{A}_n^\delta h(X_n^\delta)\} = 0$ for all functions $h: \mathbb{Z}^d \rightarrow \mathbb{R}$, where

$$\mathcal{A}_n^\delta h(X) := n \sum_{J \in \mathcal{J}} g_\delta^J(n^{-1}X) \{h(X+J) - h(X)\}, \quad X \in \mathbb{Z}^d. \quad (1.12)$$

The essence of Stein's method for total variation approximation is to find a function $h_B = h_{B,n}^\delta$ that solves the equation

$$\mathcal{A}_n^\delta h_B(X) = \mathbb{1}_B(X) - \Pi_n^\delta\{B\}, \quad X \in \tilde{B}_{n,\delta}(c), \quad (1.13)$$

for each $B \subset \tilde{B}_{n,\delta}(c)$. Then, if W is any random element of \mathbb{Z}^d and $B \subset \tilde{B}_{n,\delta}(c)$, it follows that

$$\begin{aligned} \mathbb{P}[W \in B] - \Pi_n^\delta\{B\} &= \mathbb{E}\{(\mathbb{1}_B(W) - \Pi_n^\delta\{B\})I[W \in \tilde{B}_{n,\delta'}(c)]\} \\ &\quad - \Pi_n^\delta\{B\}\mathbb{P}[W \notin \tilde{B}_{n,\delta'}(c)], \end{aligned}$$

for any $\delta' \leq \delta$, so that

$$d_{\text{TV}}(\mathcal{L}(W), \Pi_n^\delta) \leq \sup_{B \subset \tilde{B}_{n,\delta}(c)} |\mathbb{E}\{\mathcal{A}_n^\delta h_B(W)I[W \in \tilde{B}_{n,\delta'}(c)]\}| + \mathbb{P}[W \notin \tilde{B}_{n,\delta'}(c)]. \quad (1.14)$$

Showing that $\mathcal{L}(W)$ is close to Π_n^δ in total variation thus reduces to showing that the right hand side of (1.14) is small. Bounding the probability $\mathbb{P}[W \notin \tilde{B}_{n,\delta'}(c)]$ typically involves direct estimates, such as Chebyshev's inequality. Thus the main effort goes into bounding $\mathbb{E}\{\mathcal{A}_n^\delta h_B(W)\}$.

In order to extract the essential parts of $\mathbb{E}\{\mathcal{A}_n^\delta h_B(W)\}$, we expand the expression for $\mathcal{A}_n^\delta h_B(X)$, using Newton's expansion. To control the remainders in the expansion, we need to be able to control the magnitudes of the first and second differences

$$\Delta_j h_B(X) := h_B(X + e^{(j)}) - h_B(X); \quad \Delta_{jk} h_B(X) := \Delta_j(\Delta_k h_B)(X), \quad (1.15)$$

for $1 \leq j, k \leq d$, where $e^{(j)}$ denotes the j -th coordinate vector. We obtain bounds for these, given in Theorem 4.1, within a ball $|X - nc|_\Sigma \leq n\delta/4$, for δ small enough. They are derived using the explicit representation

$$h_B(X) := h_{B,n}^\delta(X) = - \int_0^\infty (\mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = X] - \Pi_n^\delta\{B\}) dt, \quad (1.16)$$

(see Kemeny & Snell (1960, Theorem 5.13(d); 1961, Equation (9))), and depend on careful analysis of the Markov process X_n^δ . This is carried out in Sections 2 and 3. For the remainders in the expansion of $\mathbb{E}\{\mathcal{A}_n^\delta h_B(W)\}$ to be small, we also need to know that $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ is small for each $1 \leq j \leq d$, and that $\mathbb{E}|W - nc|_\Sigma^2 \leq Vn$ for some constant V . This is true if $W \sim \Pi_n^\delta$, as is shown in Proposition 6.2, but needs to be proved separately for any W that is to be approximated by Π_n^δ .

As a result of these considerations, provided that $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ is small for each $1 \leq j \leq d$ and that $\mathbb{E}|W - nc|_\Sigma^2 \leq Vn$, we shall have shown, for suitable $\delta > 0$, that $\mathbb{E}\{\mathcal{A}_n^\delta h_B(W) I_n^\delta(W)\}$ is close to $\mathbb{E}\{\mathcal{A}_n h_B(W) I_n^\delta(W)\}$, for suitable $\delta > 0$, where $I_n^\delta(X) := I[|X - nc|_\Sigma \leq n\delta/3]$ and \mathcal{A}_n is as in (1.1). Hence, for any integer valued random vector W such that $\mathbb{E}\{\mathcal{A}_n h_B(W) I_n^\delta(W)\}$ is uniformly small for all $B \subset \tilde{B}_{n,\delta}(c)$, $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ is small for each $1 \leq j \leq d$ and $\mathbb{E}|W - nc|_\Sigma^2 \leq Vn$, it follows from (1.14) that $d_{\text{TV}}(\mathcal{L}(W), \Pi_n^\delta)$ is small. This conclusion is established in Theorem 4.7.

In Theorem 5.3, using the results of Theorem 4.1, we show for $W \sim \mathcal{DN}_d(nc, n\Sigma)$ that

$$|\mathbb{E}\{\mathcal{A}_n h_B(W) I_n^\delta(W)\}| = O(n^{-1/2} \log n),$$

uniformly for all B , that $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) = O(n^{-1/2})$ for each $1 \leq j \leq d$ as $n \rightarrow \infty$, and that $\mathbb{E}|W - nc|_\Sigma^2 \leq Vn$ for a constant V . It then follows from Theorem 4.7 that $d_{\text{TV}}(\mathcal{DN}_d(nc, n\Sigma), \Pi_n^\delta) = O(n^{-1/2} \log n)$ also. This means that we can use $\mathcal{DN}_d(nc, n\Sigma)$ instead of the equilibrium distribution Π_n^δ as our reference distribution, at the cost of an extra error in total variation of order $O(n^{-1/2} \log n)$, and this typically no larger than the order of the error bound for approximation using Π_n^δ , obtained in Theorem 4.7. Note also that the expectation $\mathbb{E}\{\mathcal{A}_n h_B(W) I_n^\delta(W)\}$ can be replaced by its analogue $\mathbb{E}\{\mathcal{A}_n h_B(W) I[|W - nc| \leq n\delta'/3]\}$, with truncation outside a Euclidean ball, without much extra error, provided that δ' is such that $B_{\delta'/3}(c) := \{w : |w - c| \leq \delta'/3\} \subset B_{\delta/3, \Sigma}(c)$. This follows from the assumption that $\mathbb{E}|W - nc|_\Sigma^2 \leq Vn$, and from the bounds obtained on the first and second differences of the functions h_B ; see Appendix 7.5.

We are now in a position to state our main result, Theorem 1.1. In order to do so, we need a little more notation. For a $d \times d$ positive definite symmetric matrix S , we write $\bar{\lambda}(S)$ for $d^{-1} \text{Tr}(S)$, $\lambda_{\min}(S)$ and $\lambda_{\max}(S)$ for its smallest and largest eigenvalues, respectively, and $\rho(S) := \lambda_{\max}(S)/\lambda_{\min}(S)$ for its condition number; we use $\text{Sp}'(S)$ to denote the triple $(\bar{\lambda}(S), \lambda_{\min}(S), \lambda_{\max}(S))$. Then, for a real function $h: \mathbb{Z}^d \rightarrow \mathbb{R}$, we define

$$\|\Delta h(X)\|_\infty := \max_{1 \leq i \leq d} |\Delta_i h(X)|; \quad \|\Delta^2 h(X)\|_\infty := \max_{1 \leq i, j \leq d} |\Delta_{ij} h(X)|.$$

For any $a > 0$, we also set

$$\begin{aligned}\|h\|_{a,\infty} &:= \max\{|h(X)| : X \in \mathbb{Z}^d, |X - nc| \leq a\}; \\ \|\Delta h\|_{a,\infty} &:= \max\{\|\Delta h(X)\|_\infty : X \in \mathbb{Z}^d, |X - nc| \leq a\}; \\ \|\Delta^2 h\|_{a,\infty} &:= \max\{\|\Delta^2 h(X)\|_\infty : X \in \mathbb{Z}^d, |X - nc| \leq a\},\end{aligned}\tag{1.17}$$

where nc is determined by the context.

Theorem 1.1. *Suppose that $c \in \mathbb{R}^d$, that σ^2 is a $d \times d$ positive definite symmetric matrix, and that A is a $d \times d$ matrix, all of whose eigenvalues have negative real parts; let \mathcal{A}_n be as defined in (1.1). Let Σ be the positive definite symmetric solution to the equation $A\Sigma + \Sigma A^T + \sigma^2 = 0$, and write $\bar{\Lambda} := \bar{\lambda}(\sigma^2)$, and*

$$\eta_0 := \min \left\{ \frac{4\bar{\Lambda}\sqrt{\rho(\Sigma)}}{\lambda_{\min}(\sigma^2)}, \frac{\lambda_{\min}(\sigma^2)}{24\|A\|\sqrt{\rho(\Sigma)}} \right\},$$

where $\|A\|$ denotes the spectral norm of A . Then, for any $0 < \eta \leq \eta_0$ and any $V > 0$, there exist $C_{1.1}(V, \eta), n_{1.1}(V, \eta) < \infty$ depending only on $\text{Sp}'(\sigma^2/\bar{\Lambda}), \text{Sp}'(\Sigma), \|A\|/\bar{\Lambda}, V$ and η , but not on d , with the following property: if W is any random vector in \mathbb{Z}^d such that, for some $\varepsilon_1, \varepsilon_{20}, \varepsilon_{21}$ and ε_{22} ,

- (a) $\mathbb{E}|W - nc|^2 \leq dVn$;
- (b) $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon_1$, for each $1 \leq j \leq d$;
- (c) $|\mathbb{E}\{\mathcal{A}_n h(W)\}|I[|W - nc| \leq n\eta/6]|$
 $\leq \bar{\Lambda}(\varepsilon_{20}\|h\|_{n\eta_0/4,\infty} + \varepsilon_{21}n^{1/2}\|\Delta h\|_{n\eta_0/4,\infty} + \varepsilon_{22}n\|\Delta^2 h\|_{n\eta_0/4,\infty}),$

for some $n \geq n_{1.1}(V, \eta)$, then it follows that

$$\begin{aligned}d_{\text{TV}}(\mathcal{L}(W), \mathcal{DN}_d(nc, n\Sigma)) \\ \leq C_{1.1}(V, \eta)(d^3n^{-1/2} + d^{7/2}(\bar{\gamma}(\sigma^2)/\bar{\Lambda})\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22}) \log n.\end{aligned}$$

Here, $\bar{\gamma}(\sigma^2)$ is derived from σ^2 by way of Lemma 5.4 and (5.11), and the ratio $d^{-1/2}\bar{\gamma}(\sigma^2)/\bar{\Lambda}$ is bounded by a function of $\rho(\sigma^2)$.

The estimate required in Condition (c), apart from the truncation to $|W - nc| \leq n\eta/6$, is typical of those that are needed for multivariate normal approximation using Stein's method, and Condition (a) is more or less automatic. The extra work needed, to translate multivariate normal approximation into discrete normal approximation in total variation, is that involved in establishing Condition (b). If Condition (b) is *not* satisfied, then total

variation approximation by the discrete normal cannot be accurate, since, if $W \sim \mathcal{DN}_d(nc, n\Sigma)$ and Σ is non-singular, Condition (b) is satisfied with $\varepsilon_1 = O(n^{-1/2})$: see Theorem 5.3.

Of course, for the bound to be useful, the ε -quantities must be small enough to outweigh the $\log n$ and constant factors. Under many circumstances, such as in the example in Section 6.4.1, when considering asymptotics as $n \rightarrow \infty$, they are of order $O(n^{-1/2})$, and then, for fixed d , the error bound is of order $O(n^{-1/2} \log n)$.

Note that the equilibrium distribution Π_n^δ remains the same if all the transition rates are multiplied by a common factor $a > 0$, reflecting a new choice of time scale. This has the effect of multiplying the generator \mathcal{A}_n^δ by a , whereas the solutions h_B of equation (1.13) are divided by a . The matrices A and σ^2 are also multiplied by a , leaving Σ unchanged, but multiplying the generator \mathcal{A}_n by a . A factor $\bar{\Lambda}$ is introduced on the right hand side of the inequality in Condition (c) of Theorem 1.1, in order to make the remainder of the bound invariant under such transformations, and the constant $C_{1.1}$ is then invariant also.

The quantity $\bar{\gamma}(\sigma^2)$ is rather mysterious, being defined only indirectly in terms of σ^2 . Given the matrices A and σ^2 , it is necessary to find a Markov population process with transition rates as in (1.4), for which (1.6) and (1.7) give exactly the right matrices A and σ^2 . When this has been done, we set

$$\bar{\gamma}(\sigma^2) := d^{-3/2} \sum_{J \in \mathcal{J}} |J|^3 g^J(c).$$

The detail is to be found in Section 5.2. Note that the ratio $\bar{\gamma}(\sigma^2)/\bar{\Lambda}$ is invariant under multiplication of the transition rates by a common factor a . In applications of the theorem, where W is not derived from a Markov population process, the ‘large’ parameter n also needs to be specified. Condition (a) of Theorem 1.1 suggests that it should be chosen to be comparable to $d^{-1}\mathbb{E}|W - \mathbb{E}W|^2$, as is also implied by the use of $n\Sigma$ as parameter in the approximating discrete normal distribution. This matter is discussed further in Remark 6.4, in the context of linear regression pairs.

We make some effort to make explicit the typical dependence of the error bounds on the dimension d . This is largely for comparison with the error bounds derived by Bentkus (2003) and Fang (2014) for approximation, with respect to the convex sets metric, of standardized sums of independent random vectors by the standard d -dimensional normal distribution. Here, since multiplicative standardization makes no sense in the domain of random vectors with integer coordinates, there are more quantities than just dimension that may affect the sizes of the approximation errors. Nonetheless, we attempt some comparison with the above approximations. To do so, we think

of many quantities, such as the eigenvalues of σ^2 , A and Σ , as being bounded away from zero and infinity as d varies, and the traces of these matrices thus being thought of as having order d . This is because, in the standardized setting, using the Stein approach as in Götze (1990) or Fang (2014), one has $\sigma^2 = 2I$, $A = -I$ and $\Sigma = I$. Our bounds then also involve the values of other parameters, in particular $\|A\|$ and the elements of $\text{Sp}'(\sigma^2)$ and $\text{Sp}'(\Sigma)$, in a way that can be deduced from our arguments, but that we do not attempt to make explicit, other than that their dependence on these parameters is continuous. However, we always work in terms of approximations for fixed values of n and the parameters of a problem, so that implicit orders of magnitude play no direct part in the results that we obtain.

The structure of the paper is as follows. A brief taste of the results to be obtained is given in Section 1.1. Then, in Sections 2 and 3, we establish the facts about the behaviour of the Markov population process X_n^δ that we need for solving the Stein equation derived from its generator. In Section 2.2, we define the processes X_n^δ on \mathbb{Z}^d that we use, and show that they stay concentrated at a distance of order $O(n^{1/2})$ from their centres. Using these results, we are able to show in Section 3 that the distribution of $X_n^\delta(U)$ depends smoothly on the value of its initial state $X_n^\delta(0)$, and that the initial state plays only a small part in this distribution when U is not small. In particular, two copies of X_n^δ , started in different states, can be coupled so as to coincide quite quickly. These yield the tools necessary for bounding the solutions to the Stein equation associated with X_n^δ , completed in Theorem 4.1. In the remainder of Section 4, this information is translated into an approximation theorem, Theorem 4.7, quantifying approximation by the equilibrium distribution Π_n^δ of X_n^δ , using the *approximate* generator \mathcal{A}_n as basis for the Stein argument. This is translated in Section 5 into the discrete normal approximation of Theorem 5.5, which gives three conditions to be checked in order to conclude discrete normal approximation in total variation. Its conditions make use of the $|\cdot|_\Sigma$ -metric, but the theorem is shown in Appendix 7.5 to imply Theorem 1.1, whose conditions are expressed in terms of the Euclidean metric. A number of other proofs that involve lengthy calculations are also deferred to the Appendix.

Section 6 gives some illustrations of how to apply the discrete normal approximation theorems. The first example, given in Section 6.1, concerns the equilibrium distributions of rather general Markov population processes. For the special processes of Section 3, the approximation already follows from Theorem 5.5, but more general processes require some extra argument. The next obvious choice is that of sums of independent random vectors. These are treated in Section 6.3, as an example of Theorem 6.7. This is a theorem that can be applied within the more general and widely applicable ‘linear regres-

sion pairs' structure of Section 6.2. The bounds that it gives are expressed in terms of components that frequently appear in normal approximation bounds using Stein's method. Linear regression pairs include the exchangeable pairs developed in Stein (1986), whose extra structure facilitate the checking of Condition (b) of Theorem 5.5. We conclude with an application of these results to the joint distribution of the numbers of monochrome edges in a graph colouring problem.

1.1 Applications

Theorem 1.1 is somewhat forbidding. Before we get into the detailed development, we give a rather simplified corollary of the theorem in the context of exchangeable pairs with the approximate linear regression property, and sketch an illustrative example.

Suppose that (W, W') is a pair of random integer valued d -vectors, defined on the same probability space, such that the pairs (W, W') and (W', W) have the same distribution. Assume that $\mathbb{E}\{|W|^3\} < \infty$, and write $\mu := \mathbb{E}W$. Let ξ denote the difference $W' - W$, so that $\mathbb{E}\xi = 0$, and set $\sigma^2 := \mathbb{E}\{\xi\xi^T\}$, assumed positive definite, and $\chi := \mathbb{E}\{|\xi|^3\}$. Assume that, for some $n > 0$ and for some $d \times d$ matrix A with spectral norm $\|A\|$, all of whose eigenvalues have negative real parts, we have

$$\begin{aligned}\mathbb{E}\{\xi \mid W\} &= n^{-1}A(W - \mu) + \{\|A\|/n\}^{1/2}R_1(W); \\ \sigma^2(W) &:= \mathbb{E}\{\xi\xi^T \mid W\} = \sigma^2 + R_2(W).\end{aligned}\tag{1.18}$$

Clearly, $\mathbb{E}\{R_1(W)\} = 0$ and $\mathbb{E}\{R_2(W)\} = 0$. Write $L := (\|A\|/n)^{1/2}\chi\{\text{Tr}(\sigma^2)\}^{-3/2}$, and assume that

$$\{\mathbb{E}|\Sigma^{-1/2}R_1(W)|^3\}^{1/3} \leq \frac{\lambda_{\min}(\sigma^2)}{8\lambda_{\max}(\Sigma)}\sqrt{\frac{d}{2\|A\|}},$$

where Σ is the positive definite solution of $A\Sigma + \Sigma A^T + \sigma^2 = 0$. Let \mathcal{J} be the set of d -vectors such that $q^J := \mathbb{P}[\xi = J] > 0$. Suppose that \mathcal{J} is finite, and that each of the coordinate vectors $e^{(j)}$, $1 \leq j \leq d$, can be obtained as a (finite) sum of elements of \mathcal{J} . For $Q^J(W) := \mathbb{P}[\xi = J \mid W]$, we set

$$u^J := (q^J)^{-1}\mathbb{E}|Q^J(W) - q^J|,$$

and $u^* := \max_{J \in \mathcal{J}} u^J$.

Theorem 1.2. *Under the above circumstances, there exist constants n_0 and C , depending on d, σ^2, \mathcal{J} and A , such that, if $n \geq n_0$, we have*

$$d_{TV}(\mathcal{L}(W), \mathcal{DN}_d(\mu, n\Sigma)) \leq C \log n \{L(1 + n^{1/2}u^*) + \mathbb{E}|R_1(W)|\}.$$

The key elements in the bound are L , which is the analogue of the Lyapunov ratio appearing in the Berry–Esseen error bound, u^* , which can often be shown to be small by a variance calculation, and the inaccuracy of the linear regression (1.18), expressed by $\mathbb{E}|R_1(W)|$. In examples such as the one that follows, the resulting bound is of order $O(n^{-1/2} \log n)$. The theorem can be deduced from Theorem 6.7, Lemma 6.11 and Corollary 6.12.

As an example, suppose that G_n is an r -regular graph on n vertices. Let the vertices be coloured independently, each with one of m colours, the probability of choosing colour i being $p_i > 0$, $1 \leq i \leq m$. Let N_i denote the number of vertices having colour i , and let M_i denote the number of edges joining pairs of vertices that both have colour i . We are interested in approximating the joint distribution of

$$W := (M_1, \dots, M_m, N_1, \dots, N_{m-1}) := (W_1, \dots, W_m, W_{m+1}, \dots, W_{2m-1}),$$

when n becomes large, while r , m and p_1, \dots, p_m remain fixed; the detailed structure of G_n does not appear in the approximation. Multivariate normal approximation in a smooth metric was proved by Rinott & Rotar (1996), and in the convex sets metric by Chen, Goldstein & Shao (2011, pp.333–334), both with error of order $O(n^{-1/2} \log n)$. Theorem 1.2 shows that the same order of error actually holds in total variation, provided that $m \geq 3$; the details are given in Section 6.4.1. For $m = 2$, the distribution of W is concentrated on a sub-lattice of \mathbb{Z}^3 , so that discrete normal approximation is not good (but it can be deduced for the pair (M_1, N_1)). The exchangeable pair is constructed by realizing W from a random colouring of the vertices, and then randomly re-colouring one of the vertices to give W' . The resulting regression is exact, implying that $R_1(w) = 0$ for all w . The set \mathcal{J} is fixed and finite, so that $L = O(n^{-1/2})$, and, for each J , $\mathbb{E}(Q^J(W) - q^J)^2$ can simply be shown to be of order $O(n^{-1})$ — the calculation is as for the variance of a sum of n very weakly dependent indicators. If $m \geq 3$, each coordinate vector $e^{(j)}$, $1 \leq j \leq 2m-1$, can be obtained as a sum of elements of \mathcal{J} , but this cannot be done if $m = 2$. The analogous problem, in which the proportions of vertices of each colour are held (almost) fixed, but randomly assigned to the vertices, can be treated in much the same way. The exchangeable pair is obtained by swapping the colours of two vertices, and the treatment of $\mathbb{E}(Q^J(W) - q^J)^2$ becomes a little messier.

2 The analysis of X_n^δ : general processes

2.1 Main assumptions

The main arguments of the paper are based on the analysis of a sequence of Markov population processes X_n , whose transition rates are given in (1.4). For some $\delta_0 > 0$, we make the following assumptions.

Assumption G0. The equations (1.5) have an equilibrium c ; thus $F(c) = 0$.

Assumption G1. All eigenvalues of the matrix $A := DF(c)$ have negative real parts.

Assumption G2. For each $J \in \mathcal{J}$, the function g^J is of class C^2 in the set $B_{\delta_0}(c) := \{x : |x - c| \leq \delta_0\}$.

Assumption G3. There exists $\varepsilon_0 > 0$ such that

$$\inf_{x \in B_{\delta_0}(c)} g^J(x) \geq \varepsilon_0 g^J(c) =: \mu_0^J > 0, \quad J \in \mathcal{J}.$$

Assumption G4. For each unit vector $e^{(j)} \in \mathbb{R}^d$, $1 \leq j \leq d$, there exists a finite sequence of elements $J_1^{(j)}, \dots, J_{r(j)}^{(j)}$ of \mathcal{J} such that

$$e^{(j)} = \sum_{l=1}^{r(j)} J_l^{(j)}.$$

For d -vectors, we use $|\cdot|$ to denote the Euclidean norm, $|\cdot|_1$ to denote the ℓ_1 -norm, and $|X|_\Sigma$ to denote $|\Sigma^{-1/2}X|$. For a $d \times d$ matrix B , we define the spectral norm

$$\|B\| := \sup_{y \in \mathbb{R}^d : |y|=1} |By|,$$

and use $\|B\|_1$ to denote $\sum_{i=1}^d \sum_{j=1}^d |B_{ij}|$. Note that, for any d -vector b and $d \times d$ matrix B , the inequalities

$$d^{-1}|b|_1 \leq \sqrt{d^{-1}b^T b} \quad \text{and} \quad d^{-2}\|B\|_1 \leq \sqrt{d^{-2}\text{Tr}(B^T B)} \leq \sqrt{d^{-1}\|B\|^2}$$

yield

$$|b|_1 \leq d^{1/2}|b| \quad \text{and} \quad \|B\|_1 \leq d^{3/2}\|B\|. \quad (2.1)$$

For $g: \mathbb{R}^d \rightarrow \mathbb{R}$ twice differentiable, we set

$$\|D^2g(x)\| := \limsup_{t \rightarrow 0} \sup_{y: |y|=1} t^{-1}|Dg(x+ty) - Dg(x)|.$$

We then define the quantities

$$L_0 := \max_{J \in \mathcal{J}} \frac{|g^J|_{\delta_0}}{g^J(c)}; \quad L_1 := \max_{J \in \mathcal{J}} \frac{|Dg^J|_{\delta_0}}{g^J(c)}; \quad L_2 := \max_{J \in \mathcal{J}} \frac{\|D^2g^J\|_{\delta_0}}{g^J(c)}, \quad (2.2)$$

finite in view of Assumptions G2 and G3, where $\|H\|_\delta := \sup_{x \in B_\delta(c)} \|H(x)\|$, for any vector- or matrix-valued function H and for any choice of norm $\|\cdot\|$.

We also define

$$\begin{aligned} \Lambda &:= \sum_{J \in \mathcal{J}} g^J(c) |J|^2 = \text{Tr}(\sigma^2); & \gamma &:= \sum_{J \in \mathcal{J}} g^J(c) |J|^3; \\ J_{\max} &:= \max_{J \in \mathcal{J}} |J|; & J_{\max}^\Sigma &:= \max_{J \in \mathcal{J}} |\Sigma^{-1/2} J|; \\ \sigma_\Sigma^2 &:= \Sigma^{-1/2} \sigma^2 \Sigma^{-1/2}; & \alpha_1 &:= \frac{1}{2} \lambda_{\min}(\sigma_\Sigma^2); \\ \bar{\Lambda} &:= \bar{\lambda}(\sigma^2) = d^{-1} \Lambda; & \bar{\gamma} &:= d^{-3/2} \gamma; & \mu_* &:= \min_{J \in \mathcal{J}} \mu_0^J, \end{aligned} \quad (2.3)$$

where σ^2 is defined in (1.7), and Σ in (1.3). In the sections that follow, we establish many bounds that depend on these basic parameters. They are mainly expressed as continuous functions of the elements of the set

$$\mathcal{K} := \{L_0, L_1, L_2, \varepsilon_0, \text{Sp}'(\sigma^2/\bar{\Lambda}), \text{Sp}'(\Sigma), d^{-1} J_{\max}, \|A\|/\bar{\Lambda}, \delta_0\}, \quad (2.4)$$

and, with slight abuse of notation, are said to belong to the set \mathcal{K} . If they are also continuous functions of another parameter, such as δ , they are said to belong to $\mathcal{K}(\delta)$. The factors of $1/\bar{\Lambda}$ ensure that the quantities remain invariant if all the transition rates g^J are multiplied by the same constant. In particular, constants of the form κ_i and K_i belong to \mathcal{K} , and the implied constants in any order expressions also belong to \mathcal{K} .

The d -dependence in $\bar{\Lambda}$ and $d^{-1} J_{\max}$ is put in to ensure that the quantities do not automatically have to grow with the dimension d . It is chosen in this way for the latter in view of Lemma 5.4, and for the former by comparison with $\sigma^2 = I$. In order to avoid many provisos in the bounds, we shall assume throughout that $d \leq n^{1/4}$, which is ultimately no restriction, since our bounds are typically of no use unless d is rather smaller than $n^{1/7}$.

We note two immediate consequences of Assumptions G3 and G4.

Lemma 2.1. *Assumptions G3 and G4 imply that σ^2 is positive definite, and that, for any $\delta > 0$, there exists $n_{2.1}(\delta) < \infty$ such that the process X_n^δ is irreducible on $\tilde{B}_{n,\delta}(c)$, defined in (1.11), as long as $n \geq n_{2.1}(\delta)$.*

Proof. For the first statement, if $x^T \sigma^2 x = 0$, then $x^T J = 0$ for all $J \in \mathcal{J}$, because of Assumption G3. This, from Assumption G4, implies that $x^T e^{(j)} = 0$ for all $1 \leq j \leq d$, so that $x = 0$.

For the second statement, setting $r_{\max} := \max_{1 \leq j \leq d} r(j)$, it is immediate that, under the transitions for the Markov process X_n^δ , the states X and $X \pm e^{(j)}$ communicate, for all $1 \leq j \leq d$, as long as $|X - nc|_\Sigma < n\delta - r_{\max} J_{\max}^\Sigma$. Hence, starting from an X with $|X - nc|_\Sigma \leq \max_{1 \leq j \leq d} |e^{(j)}|_\Sigma$, it follows that all states X with $|X - nc|_\Sigma < n\delta - r_{\max} J_{\max}^\Sigma$ intercommunicate.

For the remainder, we note that, because the set \mathcal{J} is finite, the infimum $\inf_{u \in \mathbb{R}^d: |u|_\Sigma=1} \min_{J \in \mathcal{J}} u^T \Sigma^{-1} J$ is attained at some u_* . Then $\min_{J \in \mathcal{J}} u_*^T \Sigma^{-1} J \geq 0$ together with $F(c) = 0$ would imply that $u_*^T \Sigma^{-1} J = 0$ for all $J \in \mathcal{J}$; and this is impossible, as argued above. Hence there exists $k_* > 0$ such that, for all u with $|u|_\Sigma = 1$, $\min_{J \in \mathcal{J}} u^T \Sigma^{-1} J < -k_*$; without loss of generality, we can also take $k_* \leq 1$.

Taking any X with $|X - nc|_\Sigma \leq n\delta$, write $X - nc = xu$, for $u \in \mathbb{R}^d$ with $|u|_\Sigma = 1$ and $x \geq 0$. Then, noting that $\sqrt{1-y} \leq 1 - y/2$ in $0 \leq y \leq 1$, we have

$$\begin{aligned} \min_{J \in \mathcal{J}} |X + J - nc|_\Sigma &= \min_{J \in \mathcal{J}} \left\{ |X - nc|_\Sigma^2 + 2(X - nc)^T \Sigma^{-1} J + |J|_\Sigma^2 \right\}^{1/2} \\ &\leq x \left\{ 1 - 2x^{-1} k_* + x^{-2} \{J_{\max}^\Sigma\}^2 \right\}^{1/2} \\ &\leq x - k_*/2, \end{aligned}$$

provided that $x \geq \max\{k_*, \{J_{\max}^\Sigma\}^2/k_*\}$. Thus each state with $|X - nc|_\Sigma \leq n\delta$ communicates with some state X' for which $|X' - nc|_\Sigma \leq |X - nc|_\Sigma - k_*/2$, and hence, repeating this step, with one such that $|X - nc|_\Sigma < n\delta - r_{\max} J_{\max}^\Sigma$. Combining these results, we see that X_n^δ is irreducible, provided that

$$n \geq n_{2.1}(\delta) := \delta^{-1} \{ (r_{\max} + 1) J_{\max}^\Sigma + \max\{k_* \{J_{\max}^\Sigma\}^2 / k_*\} \}. \quad \square$$

If Assumption G4 is *not* satisfied, then the lattice generated by the jumps in \mathcal{J} is a proper sub-lattice of \mathbb{Z}^d .

2.2 X_n^δ stays close to nc

In this section, we show that, whatever its initial value $X_n^\delta(0)$, the process X_n^δ rapidly gets close to nc . Thereafter, it remains close to nc with high probability for a very long time. To formulate our results, we define the hitting times

$$\begin{aligned} \tau_n^\delta(\eta) &:= \inf\{u \geq 0: |X_n^\delta(u) - nc|_\Sigma \geq n\eta\}; \\ \tilde{\tau}_n^\delta(\eta) &:= \inf\{u \geq 0: |X_n^\delta(u) - nc|_\Sigma \leq n\eta\}, \end{aligned} \quad (2.5)$$

for any $0 < \eta \leq \delta \leq \delta_0$.

We begin by establishing some Lyapunov–Foster–Tweedie drift conditions, showing that X_n^δ has a strong tendency to drift towards nc in the $|\cdot|_\Sigma$ norm.

Lemma 2.2. *Let X_n be a sequence of Markov population processes, whose transition rates are given in (1.4), and such that Assumptions G0–G4 are satisfied. Define*

$$\begin{aligned} h_0(X) &:= (X - nc)^T \Sigma^{-1} (X - nc) = |X - nc|_\Sigma^2; \\ h_\theta(X) &:= \exp\{n^{-1}\theta h_0(X)\}, \quad \theta > 0. \end{aligned}$$

Then there exist positive constants $K_{2.2}, \delta_{2.2}$ and θ_1 in \mathcal{K} and $\delta'_{2.2}(d) \in \mathcal{K}(d)$ such that, for any $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$ and any $X \in \tilde{B}_{n,\delta}(c)$ with $|X - nc|_\Sigma \geq K_{2.2}\sqrt{nd}$, we have

$$\mathcal{A}_n^\delta h_0(X) \leq -\alpha_1 h_0(X); \quad \mathcal{A}_n^\delta h_\theta(X) \leq -\frac{1}{2}n^{-1}\alpha_1\theta h_0(X)h_\theta(X), \quad 0 < \theta \leq \theta_1;$$

for the latter inequality, we also require that $n \geq n_{2.2} \in \mathcal{K}$. The quantities $K_{2.2}, \delta_{2.2}, \delta'_{2.2}(d)$ and θ_1 are given in (2.11), (2.13) and (2.18).

Proof. It is immediate that, for the above choice of h_0 ,

$$h_0(X + J) - h_0(X) = J^T \Sigma^{-1} (X - nc) + (X - nc)^T \Sigma^{-1} J + J^T \Sigma^{-1} J.$$

Multiplying by $ng_\delta^J(x)$, where $x := n^{-1}X$, and adding over J , we have

$$\mathcal{A}_n^\delta h_0(X) = n\{F(x)^T \Sigma^{-1} (X - nc) + (X - nc)^T \Sigma^{-1} F(x) + \text{Tr}(\Sigma^{-1} \sigma^2(x))\}, \quad (2.6)$$

as long as $|X - nc|_\Sigma < n\delta - J_{\max}^\Sigma$, where F is as defined in (1.5), and σ^2 as in (1.7). For $|X - nc|_\Sigma \geq n\delta - J_{\max}^\Sigma$, the truncation (1.10) may change this expression: see below. Now, using (2.2), for $x, y \in B_{\delta_0}(c)$, we have

$$\begin{aligned} |F(x) - F(y) - A(x - y)| &\leq \frac{1}{2} \sum_{J \in \mathcal{J}} |J| g^J(c) L_2 |x - y| \{|x - y| + 2|y - c|\} \\ &\leq \Lambda L_2 |x - y| \{|x - y| + |y - c|\}. \end{aligned} \quad (2.7)$$

Substituting (2.7), with $y = c$, into (2.6), and using (1.3), we have

$$\begin{aligned} \mathcal{A}_n^\delta h_0(X) &\leq -(X - nc)^T \Sigma^{-1} \sigma^2 \Sigma^{-1} (X - nc) + n \text{Tr}(\Sigma^{-1} \sigma^2(x)) \\ &\quad + 2\Lambda L_2 n^{-1} \|\Sigma^{-1/2}\| |X - nc|^2 |X - nc|_\Sigma. \end{aligned} \quad (2.8)$$

Using the inequalities

$$\begin{aligned} (X - nc)^T \Sigma^{-1} \sigma^2 \Sigma^{-1} (X - nc) &\geq \lambda_{\min}(\sigma_\Sigma^2) |X - nc|_\Sigma; \\ \lambda_{\min}(\Sigma) |X - nc|_\Sigma^2 &\leq |X - nc|^2 \leq \lambda_{\max}(\Sigma) |X - nc|_\Sigma^2, \end{aligned} \quad (2.9)$$

it first follows that $(X - nc)^T \Sigma^{-1} \sigma^2 \Sigma^{-1} (X - nc) \geq 2\alpha_1 |X - nc|_\Sigma^2$. Then

$$n \text{Tr}(\Sigma^{-1} \sigma^2(x)) \leq n L_0 \text{Tr}(\sigma_\Sigma^2) \leq \frac{1}{2} \alpha_1 |X - nc|_\Sigma^2 \quad (2.10)$$

if $|X - nc|_\Sigma \geq K_{2.2} \sqrt{nd}$, where

$$K_{2.2}^2 := 4L_0 \rho(\sigma^2) \rho(\Sigma), \quad (2.11)$$

since $(1/2d\alpha_1) \text{Tr}(\sigma_\Sigma^2) \leq \rho(\sigma_\Sigma^2) \leq \rho(\sigma^2) \rho(\Sigma)$. Finally,

$$2\Lambda L_2 n^{-1} \|\Sigma^{-1/2}\| |X - nc|^2 |X - nc|_\Sigma \leq \frac{1}{2} \alpha_1 |X - nc|_\Sigma^2 \quad (2.12)$$

if $|X - nc|_\Sigma \leq n \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, where

$$\delta_{2.2} := \frac{\delta_0}{\sqrt{\lambda_{\max}(\Sigma)}}; \quad \delta'_{2.2}(d) := \frac{1}{d} \frac{\alpha_1 \sqrt{\lambda_{\min}(\Sigma)}}{4\Lambda L_2 \lambda_{\max}(\Sigma)}. \quad (2.13)$$

This proves the first part of the lemma for all X such that $|X - nc|_\Sigma < n\delta - J_{\max}^\Sigma$.

If $n\delta - J_{\max}^\Sigma \leq |X - nc|_\Sigma \leq n\delta$, we may have $g^J(n^{-1}X) > g_\delta^J(n^{-1}X) = 0$ for some J . However, from the definition of h_0 , these J represent transitions for which $h_0(X + J) - h_0(X) > 0$, and replacing $g^J(n^{-1}X)$ by zero makes the value of $\mathcal{A}_n^\delta h_0(X)$ even smaller than that given in (2.6), and hence preserves the inequality (2.8).

For the second part, taking $\delta \leq \delta_{2.2}$, we note that $e^x - 1 \leq x + x^2$ in $x \leq 1$. Now, for $J_{\max}^\Sigma \leq n\delta_{2.2}$ and $|X - nc|_\Sigma \leq n\delta_{2.2}$, we have

$$\frac{\theta}{n} |h_0(X + J) - h_0(X)| \leq \frac{\theta}{n} \{2J_{\max}^\Sigma |X - nc|_\Sigma + (J_{\max}^\Sigma)^2\} \leq 3\theta J_{\max}^\Sigma \delta_{2.2},$$

and $J_{\max}^\Sigma \leq n\delta_{2.2}$ if $n \geq (d^{-1} J_{\max}^\Sigma / \delta_{2.2})^{4/3} =: n_{2.2}$, because $n \geq d^4$. Hence it follows that $n^{-1}\theta |h_0(X + J) - h_0(X)| \leq 1$ for all $X \in \tilde{B}_{n,\delta}(c)$, if $\theta \leq \theta_1$, $n \geq n_{2.2}$ and

$$\theta_1 J_{\max}^\Sigma \delta_{2.2} \leq 1/3; \quad (2.14)$$

note that then $d\theta_1 \in \mathcal{K}$. Then, for X such that $|X - nc|_\Sigma < n\delta - J_{\max}^\Sigma$, and with $x := n^{-1}X$,

$$\mathcal{A}_n^\delta h_\theta(X) = n h_\theta(X) \sum_{J \in \mathcal{J}} g^J(x) \{e^{n^{-1}\theta(h_0(X+J)-h_0(X))} - 1\}.$$

Hence, if $|X - nc|_\Sigma < n\delta - J_{\max}^\Sigma$, we have

$$\begin{aligned} n \sum_{J \in \mathcal{J}} g^J(x) \{e^{n^{-1}\theta(h_0(X+J)-h_0(X))} - 1\} \\ \leq n^{-1}\theta \mathcal{A}_n^\delta h_0(X) + n \sum_{J \in \mathcal{J}} g^J(x) n^{-2}\theta^2 |h_0(X + J) - h_0(X)|^2. \end{aligned}$$

Since

$$\begin{aligned} |h_0(X+J) - h_0(X)|^2 &\leq \{2|X - nc|_\Sigma |J|_\Sigma + |J|_\Sigma^2\}^2 \\ &\leq |J|_\Sigma^2 (8|X - nc|_\Sigma^2 + 2(J_{\max}^\Sigma)^2), \end{aligned}$$

it follows in turn that, if $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, then

$$\begin{aligned} n \sum_{J \in \mathcal{J}} g^J(x) \{e^{n^{-1}\theta(h_0(X+J) - h_0(X))} - 1\} \\ \leq -n^{-1}\theta\alpha_1 h_0(X) + 2n^{-1}\theta^2 L_0 \text{Tr}(\sigma_\Sigma^2) \{4h_0(X) + (J_{\max}^\Sigma)^2\}, \end{aligned}$$

if $\theta \leq \theta_1$. But now, if θ_1 is also chosen so that

$$8d\theta_1 L_0 \lambda_{\max}(\sigma_\Sigma^2) \leq \frac{1}{4}\alpha_1 = \frac{1}{8}\lambda_{\min}(\sigma_\Sigma^2), \quad (2.15)$$

we have $8\theta^2 L_0 \text{Tr}(\sigma_\Sigma^2) h_0(X) \leq \frac{1}{4}\theta\alpha_1 h_0(X)$, and if

$$2d\theta_1 L_0 \lambda_{\max}(\sigma_\Sigma^2) (J_{\max}^\Sigma)^2 \leq \frac{1}{4}\alpha_1 dK_{2.2}^2 = \alpha_1 dL_0 \rho(\sigma^2) \rho(\Sigma), \quad (2.16)$$

and $|X - nc|_\Sigma \geq K_{2.2}\sqrt{nd}$, we have $2\theta^2 L_0 \text{Tr}(\sigma_\Sigma^2) (J_{\max}^\Sigma)^2 \leq \frac{1}{4}\theta\alpha_1 h_0(X)$ also, so that then

$$n \sum_{J \in \mathcal{J}} g^J(x) \{e^{n^{-1}\theta(h_0(X+J) - h_0(X))} - 1\} \leq -\frac{1}{2}n^{-1}\alpha_1 \theta h_0(X). \quad (2.17)$$

Note that (2.14), (2.15) and (2.16) are satisfied by choosing

$$d\theta_1 = \min\{1/(d^{-1}J_{\max}^\Sigma \delta_{2.2}), 1/64L_0 \rho(\sigma^2) \rho(\Sigma), 1/4(d^{-1}J_{\max}^\Sigma)^2\} \in \mathcal{K}, \quad (2.18)$$

since we assume that $n \geq d^4$. As for the first part, if $n\delta - J_{\max}^\Sigma \leq |X - nc|_\Sigma \leq n\delta$, the inequality (2.17) is still true, completing the proof of the second statement of the lemma. \square

Remark 2.3. If the functions g^J are linear within $B_{\delta_0, \Sigma}$, then $L_2 = 0$, and we can take $\min\{\delta_{2.2}, \delta'_{2.2}(d)\} = \delta_{2.2} = \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$.

The first of the drift inequalities in Lemma 2.2 is now used to show that X_n^δ quickly reaches even small balls around nc , if $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$.

Lemma 2.4. Let X_n be a sequence of Markov population processes, whose transition rates are given in (1.4), and such that Assumptions G0–G4 are satisfied. Let α_1 be as in (2.3) and $K_{2.2}$, $\delta_{2.2}$ and $\delta'_{2.2}(d)$ as in Lemma 2.2. Then, if $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$ and $\eta > \max\{K_{2.2}\sqrt{d/n}, 2n^{-1}J_{\max}^\Sigma\}$, we have

$$\mathbb{P}[\tilde{\tau}_n^\delta(\eta) > t \mid X_n^\delta(0) = X_0] \leq 4(n\eta)^{-2} |X_0 - nc|_\Sigma^2 e^{-\alpha_1 t}.$$

Proof. As before, let $h_0(X) := |X - nc|_\Sigma^2$, and define $M_0(t) := h_0(X_n^\delta(t))e^{\alpha_1 t}$. Then it follows from the first part of Lemma 2.2, by a standard argument, that $M_0(t \wedge \tilde{\tau}_n^\delta(K_{2.2}\sqrt{d/n}))$, $t \geq 0$, is a non-negative supermartingale. This implies that

$$\begin{aligned} (n\eta - J_{\max}^\Sigma)^2 \mathbb{E}\{e^{\alpha_1 \tilde{\tau}_n^\delta(\eta)} \mathbb{1}\{\tilde{\tau}_n^\delta(\eta) \leq t\} \mid X_n^\delta(0) = X_0\} \\ \leq \mathbb{E}\{M_0(t \wedge \tilde{\tau}_n^\delta(\eta)) \mid X_n^\delta(0) = X_0\} \leq h_0(X_0), \end{aligned}$$

since $h_0(X_n^\delta(\tilde{\tau}_n^\delta(\eta))) \geq (n\eta - J_{\max}^\Sigma)^2$, because the jumps of X_n^δ are bounded in Σ -norm by J_{\max}^Σ . Letting $t \rightarrow \infty$, we have

$$\mathbb{E}\{e^{\alpha_1 \tilde{\tau}_n^\delta(\eta)} \mid X_n^\delta(0) = X_0\} \leq \left\{ \frac{|X_0 - nc|_\Sigma}{n\eta - J_{\max}^\Sigma} \right\}^2.$$

The lemma now follows immediately. \square

The second drift inequality in Lemma 2.2 implies that the process X_n^δ takes a long time to get far away from neighbourhoods of nc . For use in what follows, we define

$$\psi(n) := 4\sqrt{\frac{\log n}{(d\theta_1)n^{3/4}}} \quad \text{and} \quad \psi^{-1}(\eta) := \min\{n \geq 4: \psi(n) \leq \eta\}. \quad (2.19)$$

Lemma 2.5. *Let X_n be a sequence of Markov population processes, whose transition rates are given in (1.4), and such that Assumptions G0–G4 are satisfied. Then there exists $K_{2.5} \in \mathcal{K}$ such that, for all $\eta \leq \delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$ and for θ_1 as in Lemma 2.2, we have*

$$\mathbb{P}[\tau_n^\delta(\eta) \leq t \mid X_n^\delta(0) = X_0] \leq (nK_{2.5}\Lambda t + \exp\{n^{-1}\theta_1|X_0 - nc|_\Sigma^2\})e^{-n\theta_1\eta^2},$$

if $n \geq n_{2.2}$. In particular, for any $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, for any $\eta \leq \delta$, and for any $T > 0$, there exists $n_{2.5}(T) \in \mathcal{K}(\bar{\Lambda}T)$ such that, for all $|X_0 - nc|_\Sigma \leq n\eta/2$ and $t \leq T$, we have

$$\mathbb{P}[\tau_n^\delta(3\eta/4) \leq t \mid X_n^\delta(0) = X_0] \leq 2n^{-4},$$

as long as $n \geq \max\{n_{2.5}(T), \psi^{-1}(\eta)\}$. The quantities $K_{2.5}$ and $n_{2.5}(T)$ are defined in (2.21) and (2.22), respectively.

Proof. It follows from the second part of Lemma 2.2 that, for $0 \leq \theta \leq \theta_1$,

$$M_\theta(t) := h_\theta(X_n^\delta(t)) - H_\theta \int_0^t \mathbb{1}\{|X_n^\delta(s) - nc|_\Sigma \leq K_{2.2}\sqrt{nd}\} ds$$

is a supermartingale, where

$$H_\theta := \max_{X \in \mathbb{Z}^d: |X - nc|_\Sigma \leq K_{2.2} \sqrt{nd}} \mathcal{A}_n^\delta h_\theta(X).$$

Clearly, recalling $n \geq d^4$, H_θ is bounded by

$$n \sum_{J \in \mathcal{J}} \|g^J\|_{\delta_0} \exp\{n^{-1}\theta[K_{2.2}\sqrt{nd} + J_{\max}^\Sigma]^2\} \leq n\Lambda K_{2.5}, \quad (2.20)$$

for

$$K_{2.5} := L_0 \exp\{\theta_1[K_{2.2} + d^{-1}J_{\max}^\Sigma]^2\} \in \mathcal{K}. \quad (2.21)$$

By the optional stopping theorem, applied to $M_\theta(\min\{t, \tau_n^\delta(\eta)\})$, it thus follows that

$$e^{n\theta\eta^2} \mathbb{P}[\tau_n^\delta(\eta) \leq t \mid X_n^\delta(0) = X_0] - n\Lambda K_{2.5}t \leq \exp\{n^{-1}\theta|X_0 - nc|_\Sigma^2\},$$

proving the first claim. The second follows for $n \geq \max\{n_{2.5}(T), \psi^{-1}(\eta)\}$, where

$$n_{2.5}(T) := \max\{K_{2.5}\bar{\Lambda}T, n_{2.2}\}, \quad (2.22)$$

since, for such choices of n ,

$$nK_{2.5}\Lambda T \leq n^{9/4} \leq n^4 \leq e^{n\theta_1\eta^2/4}, \quad \text{and thus} \quad e^{-5n\theta_1\eta^2/16} \leq n^{-4}. \quad \square$$

3 The analysis of X_n^δ : special processes

In this section, we conduct a more detailed analysis of the Markov population processes that we use to provide approximate solutions to the simplified Stein equation (1.1). The theorems that follow are then used to bound the solution to the Stein equation (1.1) and its differences, using the representation given in (1.16); this is an essential step in proving discrete normal approximation theorems. For this, we only have to consider Markov population processes whose transition rates satisfy more restrictive conditions than Assumptions G0–G4. Since this simplifies some of the coming arguments, we stay within the more restricted context, though analogous results hold under our general assumptions. We retain Assumptions G0 and G1, replacing the remainder with the following assumptions.

Assumption S2. The set \mathcal{J} contains the vectors $\{\pm e^{(j)}, 1 \leq j \leq d\}$.

Assumption S3. The transition rates $g^J(x)$ are constant in $B_{\delta_0}(c)$, for all $J \in \mathcal{J} \setminus \{e^{(j)}, 1 \leq j \leq d\}$.

Assumption S4. For $1 \leq j \leq d$, $g^{e(j)}(x)$ is linear and satisfies $g^{e(j)}(x) \geq \frac{1}{2}g^{e(j)}(c)$ in $x \in B_{\delta_0}(c)$.

We write $g^{(j)} := g^{-e(j)}(c)$ for $1 \leq j \leq d$, and $g_* := \min_{1 \leq j \leq d} g^{(j)}$. We retain the definitions (2.2), noting that now $L_2 = 0$ and that $L_0 \leq 3/2$, and that ε_0 as defined in Assumption G3 could be taken to be $1/2$; we let $c(A) \leq d$ denote the maximal number of non-zero entries in any column of A . As observed in Remark 2.3, since $L_2 = 0$, we have

$$\min\{\delta_{2.2}, \delta'_{2.2}(d)\} = \delta_{2.2} = \delta_0 / \sqrt{\lambda_{\max}(\Sigma)}$$

for the upper bound on δ in Lemma 2.2.

From now on, we include $\bar{\Lambda}/g_*$ in the quantities belonging to \mathcal{K} , defining

$$\mathcal{K}^{(3.1)} := \mathcal{K} \cup \{\bar{\Lambda}/g_*\}; \quad (3.1)$$

$$n_{(3.1)} := \max \left\{ (5(d^{-1}J_{\max}^\Sigma) \max\{1, \sqrt{d\theta_1}\})^{8/3}, n_{2.5}(1/g_*) \right\} \in \mathcal{K}^{(3.1)}.$$

After some work, it follows from the definitions of ψ and $n_{(3.1)}$, and because $d^4 \leq n$, that $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$ implies that

$$\delta \geq 20n^{-3/4}(d^{-1}J_{\max}^\Sigma) \geq 20J_{\max}^\Sigma/n; \quad (3.2)$$

these inequalities are used later.

3.1 The dependence of $\mathcal{L}(X_n^\delta(U))$ on $X_n^\delta(0)$

We first show that the distribution $\mathcal{L}(X_n^\delta(U) | X_n^\delta(0) = X)$ does not change too much if the initial condition is slightly altered. The argument is based on that for one-dimensional processes given in Socoll & Barbour (2010). We begin by bounding differences of the form

$$\mathbb{E}\{f(X_n^\delta(U)) | X_n^\delta(0) = X - e^{(j)}\} - \mathbb{E}\{f(X_n^\delta(U)) | X_n^\delta(0) = X\},$$

and then prove a sharper bound on second differences.

Theorem 3.1. *Let X_n be a sequence of Markov population processes, whose transition rates are given in (1.4), and such that Assumptions G0, G1 and S2–S4 are satisfied. Fix any $\delta < \delta_{2.2}$. Then there are constants $(K_{3.1}^j, 1 \leq j \leq d)$ in $\mathcal{K}^{(3.1)}$, such that, for all $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$ as in (3.1),*

$$\begin{aligned} & \sup_{f: \|f\|_\infty=1} |\mathbb{E}\{f(X_n^\delta(U)) | X_n^\delta(0) = X - e^{(j)}\} - \mathbb{E}\{f(X_n^\delta(U)) | X_n^\delta(0) = X\}| \\ & \leq K_{3.1}^j n^{-1/2} \left(\frac{d\bar{\Lambda}}{g^{(j)}} \right)^{1/4} \max \left\{ 1, \frac{1}{(dg^{(j)}\bar{\Lambda})^{1/4}\sqrt{U}} \right\}, \end{aligned} \quad (3.3)$$

uniformly for all $U > 0$ and $|X - nc|_\Sigma \leq n\delta/2$.

Proof. For any $x \in \ell_1$ and any stochastic matrix P , we have $|x^T P|_1 \leq |x|_1$. Hence the quantity being bounded in (3.3) is non-increasing in U . We can thus take $U \leq 1/\sqrt{\Lambda g^{(j)}}$ in what follows, and use the bound obtained for $U = 1/\sqrt{\Lambda g^{(j)}}$ as a bound for all larger values of U . Note that, from the definition of g^J in Assumption G3, $g^{(j)} \leq \Lambda$ for all $1 \leq j \leq d$, so that $1/\sqrt{\Lambda g^{(j)}} \leq 1/g_*$.

We begin by realizing the chain X_n^δ with $X_n^\delta(0) = X_0$ in the form $X_n^\delta(u) := X_0 - e^{(j)} N_n^\delta(u) + W_n^\delta(u)$, where the bivariate chain (N_n^δ, W_n^δ) with state space $\mathbb{Z}_+ \times \mathbb{Z}^d$ starts at $(0, 0)$, and, at times u such that $|X_n^\delta(u) - nc|_\Sigma \leq n\delta - J_{\max}^\Sigma$, has transition rates given by

$$\begin{aligned} (l, W) &\rightarrow (l+1, W) \quad \text{at rate } ng^{(j)}; \\ (l, W) &\rightarrow (l, W+J) \quad \text{at rate } ng^J((X_0 - le^{(j)} + W)/n), \quad J \neq -e^{(j)} \in \mathcal{J}; \end{aligned} \tag{3.4}$$

note that the first of these transitions *reduces* the j -coordinate of X_n^δ by 1. At other values of X , it may be that $g_\delta^J(n^{-1}X)$ does not agree with $g^J(n^{-1}X)$, and so the transition rates of (N_n^δ, W_n^δ) may be different from those given in (3.4). For this reason, if the time interval $[0, U]$ is of interest, we treat any paths of X_n^δ for which $\sup_{0 \leq u \leq U} |X_n^\delta(u) - nc|_\Sigma > n\delta - 3J_{\max}^\Sigma$ separately; the factor 3 ensures that shifting a path by a vector $J' + J''$, for any $J', J'' \in \mathcal{J}$, still leaves it entirely within $\{X: |X - nc|_\Sigma \leq n\delta - J_{\max}\}$ over $[0, U]$.

Using the bivariate process, we deduce that

$$\begin{aligned} &d_{TV}\{\mathcal{L}_{X_0}(X_n^\delta(U)), \mathcal{L}_{X_0 - e^{(j)}}(X_n^\delta(U))\} \\ &= \frac{1}{2} \sum_{X \in \mathbb{Z}^d} |\mathbb{P}_{X_0}[X_n^\delta(U) = X + X_0] - \mathbb{P}_{X_0 - e^{(j)}}[X_n^\delta(U) = X + X_0]| \\ &= \frac{1}{2} \sum_{X \in \mathbb{Z}^d} \left| \sum_{l \geq 0} \mathbb{P}_{X_0}[N_n^\delta(U) = l] \mathbb{P}_{X_0}[W_n^\delta(U) = X + le^{(j)} \mid N_n^\delta(U) = l] \right. \\ &\quad \left. - \sum_{l \geq 1} \mathbb{P}_{X_0}[N_n^\delta(U) = l-1] \mathbb{P}_{X_0 - e^{(j)}}[W_n^\delta(U) = X + le^{(j)} \mid N_n^\delta(U) = l-1] \right| \\ &\leq \frac{1}{2} \sum_{X \in \mathbb{Z}^d} \sum_{l \geq 0} |\mathbb{P}_{X_0}[N_n^\delta(U) = l] - \mathbb{P}_{X_0}[N_n^\delta(U) = l-1]| q_{l-1, X_0 - e^{(j)}}^U(X + le^{(j)}) \\ &\quad + \frac{1}{2} \sum_{X \in \mathbb{Z}^d} \sum_{l \geq 1} |\mathbb{P}_{X_0}[N_n^\delta(U) = l] q_{l, X_0}^U(X + le^{(j)}) - q_{l-1, X_0 - e^{(j)}}^U(X + le^{(j)})|, \end{aligned} \tag{3.5}$$

where

$$q_{l, X}^U(W) := \mathbb{P}[W_n^\delta(U) = W \mid N_n^\delta(U) = l, X_n^\delta(0) = X]. \tag{3.6}$$

Now, from Barbour, Holst & Janson (1992, Proposition A.2.7),

$$\sum_{l \geq 0} |\text{Po}(\lambda)\{l\} - \text{Po}(\lambda)\{l-1\}| = 2 \max_{l \geq 0} \text{Po}(\lambda)\{l\} \leq \frac{1}{\sqrt{\lambda}}. \quad (3.7)$$

Hence, since N_n^δ is a Poisson process of rate $n\mu_0^J$ until the time

$$\hat{\tau}_n^\delta := \tau_n^\delta(\delta - 3n^{-1}J_{\max}^\Sigma), \quad (3.8)$$

where $\tau_n^\delta(\eta)$ is as defined in (2.5), it follows that the first term in (3.5) is bounded by

$$\mathbb{P}_{X_0}[\hat{\tau}_n^\delta \leq U] + \{ng^{(j)}U\}^{-1/2}. \quad (3.9)$$

Recall that $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$, so that, from (3.2), $\delta - 3n^{-1}J_{\max}^\Sigma > 3\delta/4$. Hence, for any $U \leq 1/\sqrt{\Lambda g^{(j)}} \leq 1/g_*$, we can use Lemma 2.5 and the definition of $\hat{\tau}_n^\delta$ to give

$$\mathbb{P}_X[\hat{\tau}_n^\delta \leq U] \leq \mathbb{P}_X[\tau_n^\delta(3\delta/4) \leq \{\Lambda g^{(j)}\}^{-1/2}] \leq 2n^{-4}, \quad (3.10)$$

uniformly in $|X - nc|_\Sigma \leq n\delta/2$. Putting this into (3.9), with $U = 1/\sqrt{\Lambda g^{(j)}}$, gives a contribution to $d_{TV}\{\mathcal{L}_{X_0}(X_n^\delta(U)), \mathcal{L}_{X_0-e^{(j)}}(X_n^\delta(U))\}$ from the first part of (3.5) of at most

$$n^{-4} + \frac{1}{2}\{ng^{(j)}U\}^{-1/2}. \quad (3.11)$$

It thus remains only to control the differences between the conditional probabilities $q_{l,X}^U(W)$ and $q_{l-1,X-e^{(j)}}^U(W)$.

To make the comparison between $q_{l,X}^U(W)$ and $q_{l-1,X-e^{(j)}}^U(W)$ for $l \geq 1$, we first condition on the whole paths of N_n^δ leading to the events $\{N_n^\delta(U) = l\}$ and $\{N_n^\delta(U) = l-1\}$, respectively, chosen to be suitably matched; we write

$$\begin{aligned} q_{l,X}^U(W) &= \frac{1}{U^l} \int_{[0,U]^l} ds_1 \dots ds_{l-1} ds^* \\ &\quad \mathbb{P}_X[W_n^\delta(U) = W \mid (N_n^\delta)^U = \nu_l(\cdot; s_1, \dots, s_{l-1}, s^*)]; \\ q_{l-1,X-e^{(j)}}^U(W) &= \frac{1}{U^l} \int_{[0,U]^l} ds_1 \dots ds_{l-1} ds^* \\ &\quad \mathbb{P}_{X-e^{(j)}}[W_n^\delta(U) = W \mid (N_n^\delta)^U = \nu_{l-1}(\cdot; s_1, \dots, s_{l-1})], \end{aligned} \quad (3.12)$$

where

$$\nu_r(u; t_1, \dots, t_r) := \sum_{i=1}^r \mathbb{1}_{[0,u]}(t_i), \quad (3.13)$$

and, for a function Y on \mathbb{R}_+ , Y^u is used to denote $(Y(s), 0 \leq s \leq u)$. Fixing $\mathbf{s}_{l-1} := (s_1, s_2, \dots, s_{l-1})$, let $\mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X}^U$ denote the distribution of $(W_n^\delta)^U$,

conditional on $(N_n^\delta)^U = \nu_l(\cdot; \mathbf{s}_{l-1}, s^*)$ and $X_n^\delta(0) = X$, and let $\mathbb{P}_{\mathbf{s}_{l-1}, X}^U$ denote the distribution conditional on $(N_n^\delta)^U = \nu_{l-1}(\cdot; \mathbf{s}_{l-1})$ and $X_n^\delta(0) = X$. Write $\hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(u, w^u)$ to denote the Radon–Nikodym derivative $d\mathbb{P}_{\mathbf{s}_{l-1}, X - e^{(j)}}^U / d\mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X}^U$ evaluated at the path w^u , for any $0 \leq u \leq U$. Then

$$\mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X}^U[W_n^\delta(U) = W] = \int_{\{w^U : w(U) = W\}} \hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(U, w^U) d\mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X}^U(w^U),$$

and hence

$$\begin{aligned} & \mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X}^U[W_n^\delta(U) = W] - \mathbb{P}_{\mathbf{s}_{l-1}, X - e^{(j)}}^U[W_n^\delta(U) = W] \\ &= \int \mathbb{1}_{\{W\}}(w(U)) \{1 - \hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(U, w^U)\} d\mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X}^U(w^U). \end{aligned} \quad (3.14)$$

Thus

$$\begin{aligned} & \sum_{W \in \mathbb{Z}^d} |q_{l, X}^U(W) - q_{l-1, X - e^{(j)}}^U(W)| \\ & \leq \frac{1}{U^l} \int_{[0, U]^l} ds_1 \dots ds_{l-1} ds^* \\ & \quad \sum_{W \in \mathbb{Z}^d} \mathbb{E}_{(\mathbf{s}_{l-1}, s^*), X}^U \left\{ |\hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(U, (W_n^\delta)^U) - 1| \mathbb{1}_{\{W\}}(W_n^\delta(U)) \right\} \quad (3.15) \\ & \leq \frac{2}{U^l} \int_{[0, U]^l} ds_1 \dots ds_{l-1} ds^* \mathbb{E}_{(\mathbf{s}_{l-1}, s^*), X}^U \left\{ [1 - \hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(U, (W_n^\delta)^U)]_+ \right\}. \end{aligned}$$

To evaluate the expectation, note that $\hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(u, (W_n^\delta)^u)$, $u \geq 0$, is a $\mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X}^U$ -martingale with expectation 1. Now, if the path w^U has r jumps of vectors J_1, \dots, J_r at times $t_1 < \dots < t_r$, write

$$x_Y(v) := n^{-1}(w(v) - e^{(j)} \nu_{l-1}(v; s_1, \dots, s_{l-1}) + Y), \quad (3.16)$$

and define

$$\hat{g}^{J'}(\cdot) := g^{J'}(\cdot), \quad J' \neq -e^{(j)}; \quad \hat{g}^{-e^{(j)}}(\cdot) := 0; \quad \hat{g}(\cdot) := \sum_{J' \in \mathcal{J}} \hat{g}^{J'}(\cdot). \quad (3.17)$$

Then, for $u \leq \hat{\tau}_n^\delta$, we have

$$\begin{aligned} & \hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(u, w^u) \\ &= \begin{cases} \exp \left(n \int_0^u \{ \hat{g}(x_{X-e^{(j)}}(v)) - \hat{g}(x_{X-e^{(j)}}(v) - e^{(j)}n^{-1}) \} dv \right) \\ \quad \prod_{\{k: 0 \leq t_k \leq u\}} \{ \hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-) - e^{(j)}n^{-1}) / \hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-)) \} \\ \hspace{15em} \text{if } u < s^*; \\ \exp \left(n \int_0^{s^*} \{ \hat{g}(x_{X-e^{(j)}}(v)) - \hat{g}(x_{X-e^{(j)}}(v) - e^{(j)}n^{-1}) \} dv \right) \\ \quad \prod_{\{k: 0 \leq t_k \leq s^*\}} \{ \hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-) - e^{(j)}n^{-1}) / \hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-)) \} \\ \hspace{15em} \text{if } u \geq s^*; \end{cases} \end{aligned} \quad (3.18)$$

after the ‘extra jump’ at s^* , the chains have come together. Note that $\hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X}^U(u, w^u)$ is absolutely continuous except for jumps at the times t_k . Then also, from Assumptions S3 and S4,

$$\frac{\hat{g}^{J'}(x - e^{(j)}n^{-1})}{\hat{g}^{J'}(x)} = 1, \quad J \notin \{e^{(1)}, e^{(2)}, \dots, e^{(d)}\},$$

and

$$\left| \frac{\hat{g}^{e^{(i)}}(x - e^{(j)}n^{-1})}{\hat{g}^{e^{(i)}}(x)} - 1 \right| \leq \frac{2\|Dg^{e^{(i)}}\|_{\delta_0}}{ng^{e^{(i)}}(c)} \leq 2L_1/n, \quad 1 \leq i \leq d, \quad (3.19)$$

uniformly in $|x - c| \leq \delta_0$. Hence, if we define the stopping time

$$\hat{\varphi}_n := \inf\{u \geq 0: \hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X_0}^U(u, (W_n^\delta)^u) \geq 2\}, \quad (3.20)$$

the jumps of the martingale $\hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X_0}^U(u, (W_n^\delta)^u)$, stopped at the time $\min(U, \hat{\tau}_n^\delta, \hat{\varphi}_n)$, are of size at most $4L_1/n$. Hence the stopped martingale has expected quadratic variation up to time u of at most

$$\int_0^u \left(\frac{4L_1}{n} \right)^2 n \sum_{J' \in \mathcal{J}} \|g^{J'}\|_{\delta_0}^2 dv \leq n^{-1}K(1/2)\Lambda u, \quad (3.21)$$

where $K(\varepsilon) := 4L_0(L_1/\varepsilon)^2 \in \mathcal{K}^{(3.1)}(\varepsilon)$. This in turn also implies that, for $0 < u \leq U$,

$$\mathbb{E}_{(\mathbf{s}_{l-1}, s^*), X_0}^U \{ (\hat{R}_{(\mathbf{s}_{l-1}, s^*), j, X_0}^U(u \wedge \hat{\tau}_n^\delta \wedge \hat{\varphi}_n, (W_n^\delta)^{u \wedge \hat{\tau}_n^\delta \wedge \hat{\varphi}_n}) - 1)^2 \} \leq n^{-1}K(1/2)\Lambda u. \quad (3.22)$$

Clearly, from (3.22) and from Kolmogorov’s inequality, once again taking $U = 1/\sqrt{\Lambda g^{(j)}}$,

$$\mathbb{P}_{(\mathbf{s}_{l-1}, s^*), X_0}^U [\hat{\varphi}_n < \min\{U, \hat{\tau}_n^\delta\}] \leq n^{-1}K(1/2)(\Lambda/g^{(j)})^{1/2}. \quad (3.23)$$

Hence, for this choice of U , from (3.22) and (3.23),

$$\begin{aligned} & \mathbb{E}_{(\mathbf{s}_{l-1}, s_*), X_0}^U \left\{ [1 - \hat{R}_{(\mathbf{s}_{l-1}, s_*), j, X_0}^U(U, (W_n^\delta)^U)]_+ \right\} \\ & \leq \min\{1, 2n^{-1/2} \sqrt{K(1/2)} (\Lambda/g^{(j)})^{1/4} + \mathbb{P}_{(\mathbf{s}_{l-1}, s_*), X_0}^U[\hat{\tau}_n^\delta < U]\}. \end{aligned} \quad (3.24)$$

In view of Lemma 2.5, the expectation of the term $\mathbb{P}_{(\mathbf{s}_{l-1}, s_*), X_0}^U[\hat{\tau}_n^\delta < U]$ is bounded by $2n^{-4}$, uniformly in $|X_0 - nc|_\Sigma \leq n\delta/2$, because $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$. Substituting this into (3.5), and using (3.15), it follows that

$$\begin{aligned} & \sum_{l \geq 1} \mathbb{P}[N_n^\delta(U) = l - 1] \sum_{W \in \mathbb{Z}^d} |q_{l, X_0}^U(W + le^{(j)}) - q_{l-1, X_0 - e^{(j)}}^U(W + le^{(j)})| \\ & \leq 2\{2n^{-1/2} \sqrt{K(1/2)} (\Lambda/g^{(j)})^{1/4} + 2\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U]\} \\ & \leq 2\{2n^{-1/2} d^{1/4} \sqrt{K(1/2)} (\bar{\Lambda}/g^{(j)})^{1/4} + 4n^{-4}\}, \end{aligned} \quad (3.25)$$

uniformly for X_0 such that $|X_0 - nc|_\Sigma \leq n\delta/2$, and for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$. Thus the contribution to $d_{TV}\{\mathcal{L}_{X_0}(X_n^\delta(U)), \mathcal{L}_{X_0 - e^{(j)}}(X_n^\delta(U))\}$ from the second part of (3.5) is at most

$$2d^{1/4} \sqrt{K(1/2)} (\bar{\Lambda}/g^{(j)})^{1/4} n^{-1/2} + 4n^{-4}, \quad (3.26)$$

and this, with (3.11), proves the theorem. \square

Remark 3.2. The bound (3.21) for the expected quadratic variation of the martingale $R_{(\mathbf{s}_{l-1}, s_*), X}^U(u, w^u)$ allows for contributions from all jumps J' of X_n^δ . However, as is clear from (3.18), if J' is such that $g^{J'}(y) = g^{J'}(y + Jn^{-1})$ for all positions y , then a jump of J' in X_n^δ does not change the value of $R_{(\mathbf{s}_{l-1}, s_*), X}^U(u, w^u)$. Since only jumps in the coordinate directions $e^{(j)}$ have rates that vary with position, only these contribute to the sum in (3.21).

Theorem 3.1 bounds differences of the form

$$\mathbb{E}\{f(X_n^\delta(U) | X_n^\delta(0) = X_0 - e^{(j)})\} - \mathbb{E}\{f(X_n^\delta(U) | X_n^\delta(0) = X_0)\},$$

showing that they are of order $O(n^{-1/2})$ uniformly in $U \geq 0$, for f such that $\|f\|_\infty \leq 1$. We now show that the corresponding second differences are of order $O(n^{-1})$.

Theorem 3.3. Let X_n be a sequence of Markov population processes, whose transition rates are given in (1.4), and such that Assumptions G0, G1 and S2–S4 are satisfied for some $\delta_0 > 0$. Fix any $\delta < \delta_{2.2}$. Then there are

constants $(K_{3.3}^{ji} \mid 1 \leq j, i \leq d)$ in $\mathcal{K}^{(3.1)}$ such that, for any function f with $\|f\|_\infty \leq 1$,

$$\begin{aligned} & \left| \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0 - e^{(j)} - e^{(i)}\} - \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0 - e^{(j)}\} \right. \\ & \quad \left. - \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0 - e^{(i)}\} + \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0\} \right| \\ & \leq K_{ji,3.3} n^{-1} \left(\frac{d\bar{\Lambda}}{g_{ij}^-} \right)^{1/2} \max \left\{ 1, \frac{1}{U \sqrt{d\bar{\Lambda} g_{ij}^+}} \right\}, \end{aligned} \quad (3.27)$$

uniformly for all $U > 0$, for $|X_0 - nc|_\Sigma \leq n\delta/4$, and for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$, where

$$g_{ij}^+ := \max\{g^{(i)}, g^{(j)}\} \quad \text{and} \quad g_{ij}^- := \min\{g^{(i)}, g^{(j)}\}.$$

Proof. As in the previous theorem, the supremum over f of the quantity being bounded in (3.27) is non-increasing in U , so that we can argue for $U \leq (\Lambda g_{ij}^+)^{-1/2} \leq 1/g_*$, and then use the bound for $U = (\Lambda g_{ij}^+)^{-1/2}$ for all larger values of U . We give the detailed argument for j and i distinct; it is almost identical if they are the same.

Much as for (3.5), we split off Poisson processes of $-e^{(j)}$ and $-e^{(i)}$ jumps. We write $X_n^\delta(u) := X_0 - e^{(j)} N_n^\delta(u) - e^{(i)} (N'_n)^\delta(u) + W_n^\delta(u)$, where the trivariate chain $(N_n^\delta, (N'_n)^\delta, W_n^\delta)$ with state space $\mathbb{Z}_+^2 \times \mathbb{Z}^d$ has transition rates

$$\begin{aligned} (l, l', W) &\rightarrow (l+1, l', W) && \text{at rate } ng^{(j)}; \\ (l, l', W) &\rightarrow (l, l'+1, W) && \text{at rate } ng^{(i)}; \\ (l, l', W) &\rightarrow (l, l', W+J) && \text{at rate } ng^J((X_0 - le^{(j)} - l'e^{(i)} + W)/n), \\ &&& J \notin \{-e^{(j)}, -e^{(i)}\}, \end{aligned} \quad (3.28)$$

up to the time $\hat{\tau}_n^\delta$, and starts at $(0, 0, 0)$. Defining

$$\begin{aligned} q_{l,l',X}^u(W) &:= \mathbb{P}_X[W_n^\delta(u) = W \mid N_n^\delta(u) = l, (N'_n)^\delta(u) = l']; \\ p_X(l, l', u) &:= \mathbb{P}_X[N_n^\delta(u) = l, (N'_n)^\delta(u) = l'], \end{aligned}$$

this allows us to deduce that

$$\begin{aligned} & \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0 - e^{(j)} - e^{(i)}\} - \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0 - e^{(j)}\} \\ & \quad - \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0 - e^{(i)}\} + \mathbb{E}\{f(X_n^\delta(U)) \mid X_n^\delta(0) = X_0\} \\ & = \sum_{X \in \mathbb{Z}^d} f(X) \sum_{l \geq 0} \sum_{l' \geq 0} \left\{ p_{X_0}(l-1, l'-1, U) q_{l-1, l'-1, X_0 - e^{(j)} - e^{(i)}}^U(X + le^{(j)} + l'e^{(i)}) \right. \\ & \quad - p_{X_0}(l-1, l', U) q_{l-1, l', X_0 - e^{(j)}}^U(X + le^{(j)} + l'e^{(i)}) \\ & \quad - p_{X_0}(l, l'-1, U) q_{l, l'-1, X_0 - e^{(i)}}^U(X + le^{(j)} + l'e^{(i)}) \\ & \quad \left. + p_{X_0}(l, l', U) q_{l, l', X_0}^U(X + le^{(j)} + l'e^{(i)}) \right\}. \end{aligned} \quad (3.29)$$

Write $r_{j,k,X}(l, l', u) := p_X(l - j, l' - k, u)/p_X(l, l', u)$ for $j, k \in \{0, 1\}$, and

$$R_{j,k,Y;l,l',X}^u(W) := q_{l-j,l'-k,X+Y}^u(W)/q_{l,l',X}^u(W).$$

Then the right hand side of (3.29) can be expressed as

$$\begin{aligned} & \sum_{l \geq 0} \sum_{l' \geq 0} p_{X_0}(l, l', U) \sum_{w \in \mathbb{Z}^d} f(w - le^{(j)} - l'e^{(i)}) q_{l,l',X_0}^U(w) \\ & \quad \left\{ r_{11,X_0}(l, l', U) R_{1,1,-e^{(j)}-e^{(i)};l,l',X_0}^U(w) - r_{10,X_0}(l, l', U) R_{1,0,-e^{(j)};l,l',X_0}^U(w) \right. \\ & \quad \left. - r_{01,X_0}(l, l', U) R_{0,1,-e^{(i)};l,l',X_0}^U(w) + 1 \right\}. \end{aligned} \quad (3.30)$$

We now use the decomposition

$$rR = (r-1)(R-1) + (r-1) + (R-1) + 1$$

in each term of (3.30). The sum corresponding to taking 1 yields nothing. Then, for the sum corresponding to taking $(r-1)$ alone, summing over w first and using $\|f\|_\infty \leq 1$, we have

$$\begin{aligned} & \sum_{l,l' \geq 0} p_{X_0}(l, l', U) \sum_{w \in \mathbb{Z}^d} |f(w - le^{(j)} - l'e^{(i)})| q_{l,l',X_0}^U(w) \\ & \quad |r_{11,X_0}(l, l', U) - r_{10,X_0}(l, l', U) - r_{01,X_0}(l, l', U) + 1| \quad (3.31) \\ & \leq \sum_{l \geq 0} \sum_{l' \geq 0} p_{X_0}(l, l', U) |r_{11,X_0}(l, l', U) - r_{10,X_0}(l, l', U) - r_{01,X_0}(l, l', U) + 1|. \end{aligned}$$

As for (3.5) and (3.11), the processes $(N_n^\delta, (N')_n^\delta)$ can be coupled to independent Poisson processes with rates $ng^{(j)}$ and $ng^{(i)}$ respectively on the interval $[0, U]$, with failure probability at most $\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U]$. Hence, using $\pi^{(j)}$ to denote $\text{Po}(nUg^{(j)})$, (3.31) gives a contribution to (3.30) of at most

$$\begin{aligned} & \sum_{l \geq 0} \sum_{l' \geq 0} |\pi^{(j)}\{l\} - \pi^{(j)}\{l-1\}| |\pi^{(i)}\{l'\} - \pi^{(i)}\{l'-1\}| + 4\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U] \\ & = 4d_{\text{TV}}(\pi^{(j)}, \pi^{(j)} * \varepsilon_1) d_{\text{TV}}(\pi^{(i)}, \pi^{(i)} * \varepsilon_1) + 4\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U] \\ & \leq \frac{4}{\sqrt{g^{(j)}g^{(i)}}} \frac{1}{nU} + 8n^{-4}, \end{aligned} \quad (3.32)$$

for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$, uniformly in $|X_0 - nc|_\Sigma \leq n\delta/4$.

We separate the sum corresponding to $(r-1)(R-1)$ in (3.30) into three pieces, corresponding to the subscripts $(1, 1)$, $(1, 0)$ and $(0, 1)$, and use $\|f\|_\infty \leq 1$. We then use an argument similar to that leading to (3.25);

we sketch it for the $(1, 1)$ case. First, by conditioning on the paths of N_n^δ and $(N')_n^\delta$ and using (3.40) below, it follows, much as for (3.25) and for (3.24), that, for each $l, l' \geq 0$,

$$\begin{aligned} & \sum_{w \in \mathbb{Z}^d} q_{l, l', X_0}^U(w) |1 - R_{1, 1, -e^{(j)} - e^{(i)}; l, l', X_0}^U(w)| \\ & \leq \min\{2, 2n^{-1/2} \sqrt{K(1/2)\Lambda U} + 4n^{-1} K(1/2)\Lambda U \\ & \quad + 2\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U \mid N_n^\delta(u) = l, (N')_n^\delta(u) = l']\} \\ & \leq 4n^{-1/2} \sqrt{K(1/2)\Lambda U} + 2\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U \mid N_n^\delta(u) = l, (N')_n^\delta(u) = l']. \end{aligned} \quad (3.33)$$

Then, as in treating (3.31), and using Lemma 2.5, we have

$$\begin{aligned} & \sum_{l, l' \geq 0} p_{X_0}(l, l', U) |r_{11, X_0}(l, l', U) - 1| \\ & \leq 2\{d_{\text{TV}}(\pi^{(j)}, \pi^{(j)} * \varepsilon_1) + d_{\text{TV}}(\pi^{(i)}, \pi^{(i)} * \varepsilon_1)\} + 4\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U] \\ & \leq \frac{2}{\sqrt{nUg^{(j)}}} + \frac{2}{\sqrt{nUg^{(i)}}} + 8n^{-4} \leq \frac{4}{\sqrt{nUg_{ij}^-}} + 8n^{-4}, \end{aligned} \quad (3.34)$$

for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$, uniformly in $|X_0 - nc|_\Sigma \leq n\delta/4$.

Combining the first part of (3.33) with (3.34) gives a contribution to (3.30) bounded by

$$Kn^{-1}d^{1/2}(\bar{\Lambda}/g_{ij}^-)^{1/2} + 12n^{-2}, \quad (3.35)$$

uniformly for $U \leq 1/g_*$ and $|X_0 - nc|_\Sigma \leq n\delta/4$, for $K := 4\sqrt{K(1/2)} \in \mathcal{K}^{(3.1)}$. Taking the second part of (3.33) with (3.34), it is immediate that

$$\begin{aligned} & 2 \sum_{l, l' \geq 0} p_{X_0}(l, l', U) |r_{11, X_0}(l, l', U) - 1| \mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U \mid N_n^\delta(u) = l, (N')_n^\delta(u) = l'] \\ & \leq 2n^2 \mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U] \leq 4n^{-2}, \end{aligned} \quad \mathbb{1}\{r_{11, X_0}(l, l', U) \leq n^2\}$$

by Lemma 2.5, since $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$. For the remainder, we have at most

$$\begin{aligned} & 2 \sum_{l, l' \geq 0} p_{X_0}(l, l', U) \mathbb{1}\{r_{11, X_0}(l+1, l'+1, U) > n^2\} \\ & \leq 2\mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U] + 2 \sum_{l, l' \geq 0} \pi^{(j)}\{l\} \pi^{(i)}\{l'\} \mathbb{1}\{r_{11, X_0}(l+1, l'+1, U) > n^2\}. \end{aligned}$$

Now, since

$$|p_{X_0}(l, l', U) - \pi^{(j)}\{l\} \pi^{(i)}\{l'\}| \leq \mathbb{P}_{X_0}[\hat{\tau}_n^\delta < U] \leq 2n^{-4},$$

it follows that, if

$$\min(\pi^{(j)}\{l\}, \pi^{(i)}\{l'\}, \pi^{(j)}\{l+1\}, \pi^{(i)}\{l'+1\}) \geq 2n^{-2}, \quad (3.36)$$

then

$$r_{11, X_0}(l+1, l'+1, U) \leq 3 \frac{\pi^{(j)}\{l\} \pi^{(i)}\{l'\}}{\pi^{(j)}\{l+1\} \pi^{(i)}\{l'+1\}} \leq 3 \frac{(l+1)(l'+1)}{n^2 U^2 g_{ij}^- g_{ij}^+}.$$

Then, by Proposition A.2.3 (i) of Barbour, Holst & Janson (1992), if (3.36) holds,

$$3 \frac{(l+1)(l'+1)}{n^2 U^2 g_{ij}^- g_{ij}^+} \leq 100(\log n)^2 < n^2,$$

for all $n \geq 40$; in proving the first inequality, note that we can assume that $nUg_{ij}^- \geq 1$, since the inequality in the statement of the theorem is immediate for smaller nU . This leaves only a contribution from l, l' for which (3.36) does not hold, which is at most

$$2 \sum_{l \geq 0} \{ \pi^{(j)}\{l\} \mathbb{1}\{\pi^{(j)}\{l\} \leq 2n^{-2}\} + \pi^{(i)}\{l\} \mathbb{1}\{\pi^{(i)}\{l\} \leq 2n^{-2}\} \} \leq Kn^{-3/2},$$

by Proposition A.2.3 (ii), (iii) and (iv) of Barbour, Holst & Janson (1992), for a universal constant K , because we also have $nUg_{ij}^+ \leq n$.

The trickiest sum is that corresponding to $(R-1)$ alone. Using $\|f\|_\infty \leq 1$, we need first to examine the quantity

$$\sum_{w \in \mathbb{Z}^d} q_{l, l', X_0}^U(w) \left| R_{1, 1, -e^{(j)} - e^{(i)}; l, l', X_0}^U(w) - R_{1, 0, -e^{(j)}; l, l', X_0}^U(w) - R_{0, 1, -e^{(i)}; l, l', X_0}^U(w) + 1 \right|. \quad (3.37)$$

We treat it, after conditioning on realizations of the underlying Poisson processes N_n^δ and $(N')_n^\delta$, as the expectation of the absolute value of a martingale $M^{(2)}(W_n^\delta)$ at time U . Let $W^u := (W(t), 0 \leq t \leq u)$ denote the restriction of a function W on \mathbb{R}_+ to $[0, u]$. Write $\mathbf{s}_l := (s_1, \dots, s_l)$, $\mathbf{s}'_{l'} := (s'_1, \dots, s'_{l'})$. If realizations of N_n^δ and $(N')_n^\delta$, having l and l' points respectively in $[0, U]$, are denoted by $\nu_l(\cdot; \mathbf{s}_l)$ and $\nu'_{l'}(\cdot; \mathbf{s}'_{l'})$, as in (3.13), we then denote conditional probability and expectation, given $(N_n^\delta)^U = \nu_l(\cdot; \mathbf{s}_l)$, $((N')_n^\delta)^U = \nu'_{l'}(\cdot; \mathbf{s}'_{l'})$ and $X_n^\delta(0) = X$, by $\mathbb{P}_{\mathbf{s}_l, \mathbf{s}'_{l'}, X}^U$ and $\mathbb{E}_{\mathbf{s}_l, \mathbf{s}'_{l'}, X}^U$, and we denote the corresponding conditional density of $(W_n^\delta)^u$ at the path segment W^u , with respect to some suitable reference measure, by

$$q^U(u, W^u; \mathbf{s}_l, \mathbf{s}'_{l'}, X).$$

We then define the Radon–Nikodym derivatives

$$\begin{aligned}
R_{11}^U(u, W^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) &:= \frac{q^U(u, W^u; \mathbf{s}_{l-1}, \mathbf{s}'_{l'-1}, X_0 - e^{(j)} - e^{(i)})}{q^U(u, W^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0)}; \\
R_{10}^U(u, W^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) &:= \frac{q^U(u, W^u; \mathbf{s}_{l-1}, (\mathbf{s}'_{l'-1}, s'_*), X_0 - e^{(j)})}{q^U(u, W^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0)}; \\
R_{01}^U(u, W^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) &:= \frac{q^U(u, W^u; (\mathbf{s}_{l-1}, s_*), \mathbf{s}'_{l'-1}, X_0 - e^{(i)})}{q^U(u, W^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0)};
\end{aligned}$$

these have explicit formulae analogous to (3.18). We use them to formulate the analogue of the argument used in the proof of Theorem 3.1. For example, we can write

$$\begin{aligned}
&\sum_{w \in \mathbb{Z}^d} q_{l, l', X_0}^U(w) R_{1, 1, -e^{(j)} - e^{(i)}; l, l', X_0}^U(w) \\
&= \frac{1}{U^{l+l'}} \int_{[0, U]^{l+l'}} ds_1 \dots ds_{l-1} ds_* ds'_1 \dots ds'_{l'-1} ds'_* \\
&\quad \sum_{w \in \mathbb{Z}^d} \mathbb{P}_{\mathbf{s}_{l-1}, \mathbf{s}'_{l'-1}, X_0 - e^{(j)} - e^{(i)}}^U[W(U) = w] \\
&= \frac{1}{U^{l+l'}} \int_{[0, U]^{l+l'}} ds_1 \dots ds_{l-1} ds_* ds'_1 \dots ds'_{l'-1} ds'_* \\
&\quad \sum_{w \in \mathbb{Z}^d} \mathbb{E}_{(\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0}^U \{R_{11}^U(U, W^U; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) I[W(U) = w]\}.
\end{aligned}$$

The mean zero martingale $M^{(2)}$ of main interest to us can then be expressed as

$$\begin{aligned}
M^{(2)}(W_n^\delta)(u) &:= R_{11}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) \\
&\quad - R_{10}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) \\
&\quad - R_{01}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) + 1,
\end{aligned} \tag{3.38}$$

with $(W_n^\delta)^U$ a random element with distribution $\mathbb{P}_{(\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0}^U$. We also define the martingale

$$M^{(1)}(W_n^\delta)(u) := R_{11}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) - 1,$$

for use in the proof below, as well as for the proof of the estimate of the $(1, 1)$ term in (3.33) above.

We now set $x_n^\delta(u) := n^{-1}(W_n^\delta(u) + X_0 - e^{(j)})\nu_{l-1}(u; \mathbf{s}_{l-1}) - e^{(i)}\nu'_{l'-1}(u; \mathbf{s}'_{l'-1})$ for $u < \min\{s_*, s'_*\}$. If, for $u < \min\{s_*, s'_*\}$ and $|x_n^\delta(u) - c|_\Sigma \leq \delta - 3n^{-1}J_{\max}^\Sigma$,

there is a jump of $e^{(r)}$ in W_n^δ at time u , for some $1 \leq r \leq d$, this gives rise to a jump in the martingale $M^{(2)}(W_n^\delta)$ at u of

$$\begin{aligned} & R_{11}^U(u-, (W_n^\delta)^{u-}; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l-1}, s'_*), X_0) \left(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u-) - n^{-1}(e^{(j)} + e^{(i)}))}{\hat{g}^{e^{(r)}}(x_n^\delta(u-))} - 1 \right) \\ & - R_{10}^U(u-, (W_n^\delta)^{u-}; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l-1}, s'_*), X_0) \left(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u-) - n^{-1}e^{(j)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u-))} - 1 \right) \\ & - R_{01}^U(u-, (W_n^\delta)^{u-}; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l-1}, s'_*), X_0) \left(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u-) - n^{-1}e^{(i)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u-))} - 1 \right). \end{aligned}$$

If $s_* < u < s'_*$, the elements $-n^{-1}e^{(j)}$ are removed from the arguments of $\hat{g}^{e^{(r)}}$, simplifying the considerations, but then $x_n^\delta(u)$ is replaced by $x_n^\delta(u) - n^{-1}e^{(j)}$; the elements $-n^{-1}e^{(i)}$ are removed if $s'_* < u < s_*$, and then $x_n^\delta(u)$ is replaced by $x_n^\delta(u) - n^{-1}e^{(i)}$; if $u > \max\{s_*, s'_*\}$, both elements $-n^{-1}e^{(j)}$ and $-n^{-1}e^{(i)}$ are removed, and so there is no jump. Now, because the transition rate $g^{e^{(r)}}(x)$ is linear in x ,

$$\begin{aligned} & \left(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) - n^{-1}(e^{(j)} + e^{(i)}))}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1 \right) - \left(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) - n^{-1}e^{(j)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1 \right) \\ & - \left(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) - n^{-1}e^{(i)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1 \right) = 0, \end{aligned}$$

and so R^U can be replaced by $|R^U - 1|$ when bounding the sizes of the jumps, irrespective of the relative positions of s_* , s'_* and u . Since also, from (2.2) and Assumption S4,

$$\left| \frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) + n^{-1}Y)}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1 \right| \leq 2n^{-1}|Y|L_1,$$

the remaining contributions to the jump in $M^{(2)}$ are at most

$$\begin{aligned} & \frac{4L_1}{n} \{ |R_{11}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l-1}, s'_*), X_0) - 1| \\ & + |R_{10}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l-1}, s'_*), X_0) - 1| \\ & + |R_{01}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l-1}, s'_*), X_0) - 1| \}. \end{aligned} \quad (3.39)$$

We can now bound the quadratic variation arising from each of the three terms individually, by the argument leading to (3.22). Defining

$$\tilde{\varphi}_n := \inf\{u \geq 0: \tilde{m}(u) \geq 2\},$$

where

$$\begin{aligned}\tilde{m}(u) &:= \max\{R_{11}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0), \\ &\quad R_{10}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0), \\ &\quad R_{01}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0)\},\end{aligned}$$

we use the martingale $M^{(1)}$ to give

$$\begin{aligned}\mathbb{E}_{(\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0}^U \{[R_{11}^U(u \wedge \hat{\tau}_n^\delta \wedge \tilde{\varphi}_n, (W_n^\delta)^{u \wedge \hat{\tau}_n^\delta \wedge \tilde{\varphi}_n}; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) - 1]^2\} \\ \leq n^{-1} 4K(1/2)\Lambda u,\end{aligned}\tag{3.40}$$

for $K(\varepsilon)$ as defined following (3.21); the same bound, but without the factor 4, holds for R_{10}^U and R_{01}^U also. Hence the expected quadratic variation of the martingale $M^{(2)}(W_n^\delta)$ stopped at $u \wedge \hat{\tau}_n^\delta \wedge \tilde{\varphi}_n$ is at most

$$\begin{aligned}n\Lambda \int_0^u \left(\frac{12L_1}{n}\right)^2 \left(\frac{4K(1/2)\Lambda v}{n}\right) dv \\ \leq 2n^{-2}(\Lambda u)^2(12L_1)^2 K(1/2) \leq n^{-2}K_8(\Lambda u)^2,\end{aligned}$$

uniformly in $|X_0 - nc|_\Sigma \leq n\delta$, and in $l, l', \mathbf{s}_{l-1}, \mathbf{s}'_{l'-1}, s_*$ and s'_* , for $K_8 := 2(12L_1)^2 K(1/2) \in \mathcal{K}^{(3.1)}$. This gives a contribution of at most $n^{-1}d\sqrt{K_8}(\bar{\Lambda}U)$ to (3.37), and hence to (3.30), from the expectation of $|M^{(2)} - 1|$, stopped at $U \wedge \hat{\tau}_n^\delta \wedge \tilde{\varphi}_n$.

Because the martingale $M^{(2)}(W_n^\delta)$ is not uniformly bounded from below, we can no longer use an argument as for (3.24) to bound the contributions to (3.30) from the events $\hat{\tau}_n^\delta < U$ and $\tilde{\varphi}_n < U$. Instead, we consider their contributions for each element of $M^{(2)}(W_n^\delta)$ separately. For example, writing

$$\begin{aligned}\tilde{R}_*^U &:= R_{11}^U((W_n^\delta)^U; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0); \\ \mathbb{E}_*^U &:= \mathbb{E}_{(\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0}^U; \quad \mathbb{P}_*^U := \mathbb{P}_{(\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0}^U,\end{aligned}$$

we have

$$\begin{aligned}\sum_{w \in \mathbb{Z}^d} f(w - le^{(j)} - l'e^{(i)}) \mathbb{E}_*^U \{\tilde{R}_*^U I[W_n^\delta(U) = w, \hat{\tau}_n^\delta < U]\} \\ \leq \mathbb{E}_*^U \left\{ \tilde{R}_*^U I \left[\sup_{0 \leq u \leq U} |X_0 + W_n^\delta(u) - e^{(j)}\nu(u; \mathbf{s}_{l-1}, s_*) \right. \right. \\ \left. \left. - e^{(i)}\nu'(u; \mathbf{s}'_{l'-1}, s'_*) - nc| \geq n\delta - 3J_{\max}^\Sigma \right] \right\} \\ = \mathbb{E}_{\mathbf{s}_{l-1}, \mathbf{s}'_{l'-1}, X_0 - e^{(j)} - e^{(i)}}^U \left\{ I \left[\sup_{0 \leq u \leq U} |X_0 + W_n^\delta(u) - e^{(j)}\nu(u; \mathbf{s}_{l-1}) \right. \right. \\ \left. \left. - e^{(i)}\nu'(u; \mathbf{s}'_{l'-1}) - e^{(j)}\mathbb{1}_{[s_*, U]}(u) - e^{(i)}\mathbb{1}_{[s'_*, U]}(u) - nc| \geq n\delta - 3J_{\max}^\Sigma \right] \right\} \\ \leq \mathbb{P}_{\mathbf{s}_{l-1}, \mathbf{s}'_{l'-1}, X_0 - e^{(j)} - e^{(i)}}^U [\tau_n^\delta(\delta - 3n^{-1}J_{\max}^\Sigma - 2n^{-1}/\sqrt{\lambda_{\min}(\Sigma)}) < U].\end{aligned}$$

Now both the inequalities

$$|X_0 - e^{(j)} - e^{(i)} - nc|_\Sigma \leq n\delta/2$$

and

$$\tau_n^\delta(\delta - 3n^{-1}J_{\max}^\Sigma - 2n^{-1}/\sqrt{\lambda_{\min}(\Sigma)}) \geq \tau_n^\delta(3\delta/4)$$

are satisfied if $n^{3/4}\delta > 20d^{-1}J_{\max}^\Sigma$ and $|X_0 - nc|_\Sigma \leq n\delta/4$. Taking expectations over the realizations of $(N_n^\delta)^U$ and $((N')_n^\delta)^U$ and invoking Lemma 2.5 thus gives a contribution to (3.30) of at most $2n^{-4}$, uniformly in $|X_0 - nc|_\Sigma \leq n\delta/4$, for $n \geq \max\{n_{2.5}(1/g_*), (20d^{-1}J_{\max}^\Sigma/\delta)^{4/3}, \psi^{-1}(\delta)\}$; this inequality is satisfied if $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$. Then

$$\begin{aligned} & \sum_{w \in \mathbb{Z}^d} f(w - le^{(j)} - l'e^{(i)}) \mathbb{E}_*^U \left\{ \tilde{R}_*^U I[W_n^\delta(U) = w, \tilde{\varphi}_n < \min\{U, \hat{\tau}_n^\delta\}] \right\} \\ & \leq \mathbb{E}_*^U \{ \tilde{R}_*^U I[\tilde{\varphi}_n < \min\{U, \hat{\tau}_n^\delta\}] \} \\ & \leq 2\mathbb{P}_*^U[\tilde{\varphi}_n < \min\{U, \hat{\tau}_n^\delta\}] + \mathbb{E}_*^U \{ \tilde{R}_*^U I[\tilde{\varphi}_n^{11} < \min\{U, \hat{\tau}_n^\delta\}] \}, \end{aligned}$$

where

$$\tilde{\varphi}_n^{11} := \inf\{u \geq 0: R_{11}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) \geq 2\}.$$

The first of these terms is at most $6n^{-1}K(1/2)\Lambda U$, using (3.40) and its analogues for the quantities R_{11}^U , R_{10}^U and R_{01}^U appearing in the definition of $\tilde{m}(U)$, and then applying Kolmogorov's inequality; the argument is much as for (3.23). The second is no larger than

$$2\mathbb{P}_{\mathbf{s}_{l-1}, \mathbf{s}'_{l'-1}, X_0 - e^{(j)} - e^{(i)}}^U[\tilde{\varphi}_n^{11} < \min\{U, \hat{\tau}_n^\delta\}]. \quad (3.41)$$

However, under $\mathbb{P}_{\mathbf{s}_{l-1}, \mathbf{s}'_{l'-1}, X_0 - e^{(j)} - e^{(i)}}^U$, the process

$$M' := \{1/R_{11}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0), u \geq 0\}$$

is a martingale with mean 1. Arguing much as for (3.21), its expected quadratic variation up to the time $\min\{U, \hat{\tau}_n^\delta, (\tilde{\varphi}'_n)^{11}\}$ can be shown to be at most $4n^{-1}K(1/2)\Lambda U$; here, $(\tilde{\varphi}'_n)^{11} := \inf\{u \geq 0: M'(u) \geq 2\}$. Using an argument much as that for (3.23), Kolmogorov's inequality now shows that the quantity in (3.41) is itself at most $8n^{-1}K(1/2)\Lambda U$, giving a contribution to (3.30) of order $O(n^{-1}d\Lambda U)$. Combining these considerations with (3.32) and (3.35), the inequality of the theorem follows for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta)\}$. \square

Remark 3.4. *The discussion in Remark 3.2 applies also to Theorem 3.3. It shows in particular that the factors $d\bar{\Lambda}$ appearing twice in each bound can be replaced by $c(A)g^*$, where $c(A)$ is as defined following Assumption S4, and $g^* := \max_{1 \leq j \leq d} g^{e(j)}(c)$. This is because the sum in (3.21) can be replaced by one consisting of only $c(A)$ elements, for each of which $|g^{j'}|_\delta \leq L_0 g^*$.*

3.2 Coupling copies of X_n^δ

In this section, we show that copies of X_n^δ with different initial states can be defined on the same probability space, in such a way that they coincide rather quickly. As a consequence, the total variation distance between their distributions becomes small as time increases. Our arguments are reminiscent of those in Roberts & Rosenthal (1996).

The basic coupling that we use relies mainly on the drift towards nc to achieve this. We define the process $(X_{n,1}^\delta(t), X_{n,2}^\delta(t))$ on $\tilde{B}_{n,\delta}(c) \times \tilde{B}_{n,\delta}(c)$ to have the transition rates

$$\begin{aligned} (X_1, X_2) &\rightarrow (X_1 + J, X_2 + J) & \text{at rate } &n\{g_\delta^J(n^{-1}X_1) \wedge g_\delta^J(n^{-1}X_2)\}; \\ (X_1, X_2) &\rightarrow (X_1, X_2 + J) & \text{at rate } &n\{g_\delta^J(n^{-1}X_2) - g_\delta^J(n^{-1}X_1)\}_+; \\ (X_1, X_2) &\rightarrow (X_1 + J, X_2) & \text{at rate } &n\{g_\delta^J(n^{-1}X_1) - g_\delta^J(n^{-1}X_2)\}_+, \end{aligned}$$

for each $J \in \mathcal{J}$. Let its generator be denoted by $\tilde{\mathcal{A}}_n^\delta$. Our coupling argument begins with a drift inequality.

Lemma 3.5. *Let X_n be a sequence of Markov population processes, whose transition rates are given in (1.4), and such that Assumptions G0, G1 and S2–S4 are satisfied. Define $h_1(X_1, X_2) := |X_1 - X_2|_\Sigma^2$; let α_1 be as in (2.3) and $\delta_1 := \delta_0 / \sqrt{\lambda_{\max}(\Sigma)}$. Then, for $\delta \leq \delta_1/3$, there exists $K_{3.5} \in \mathcal{K}^{(3.1)}$, defined in (3.42), such that, for all (X_1, X_2) with $\max\{|X_1 - nc|_\Sigma, |X_2 - nc|_\Sigma\} \leq n\delta - J_{\max}^\Sigma$ and $|X_1 - X_2|_\Sigma \geq dK_{3.5}$, we have*

$$\tilde{\mathcal{A}}_n^\delta h_2(X_1, X_2) \leq -\frac{1}{2}\alpha_1 h_2(X_1, X_2),$$

where $h_2(X_1, X_2) := h_1(X_1, X_2) + d^2 K_{3.5}^2$.

Proof. For any $\xi > 0$, write $h^{(\xi)}(X_1, X_2) := h_1(X_1, X_2) + \xi$. By the definition of h_1 , the transitions where both components of (X_1, X_2) make the same jump make no contribution to $\tilde{\mathcal{A}}_n^\delta h^{(\xi)}(X_1, X_2)$. Hence, for (X_1, X_2) with $\max\{|X_1 - nc|_\Sigma, |X_2 - nc|_\Sigma\} \leq n\delta - J_{\max}^\Sigma$, and writing $x_i := n^{-1}X_i$, $i = 1, 2$,

we have

$$\begin{aligned}
\tilde{\mathcal{A}}_n^\delta h^{(\xi)}(X_1, X_2) &= n \sum_{J \in \mathcal{J}} \left\{ (g^J(x_1) - g^J(x_2))_+ \{2J^T \Sigma^{-1}(X_1 - X_2) + J^T \Sigma^{-1} J\} \right. \\
&\quad \left. + (g^J(x_2) - g^J(x_1))_+ \{-2J^T \Sigma^{-1}(X_1 - X_2) + J^T \Sigma^{-1} J\} \right\} \\
&= 2n(F(x_1) - F(x_2))^T \Sigma^{-1}(X_1 - X_2) + n \sum_{J \in \mathcal{J}} |g^J(x_1) - g^J(x_2)| |J|_\Sigma^2,
\end{aligned}$$

since, for such (X_1, X_2) , $g_\delta^J(x_1) = g^J(x_1)$ and $g_\delta^J(x_2) = g^J(x_2)$. Now, since the transition rates $g^J(x)$ are all linear in x , we have

$$\begin{aligned}
2n(F(x_1) - F(x_2))^T \Sigma^{-1}(X_1 - X_2) &= -(X_1 - X_2)^T \Sigma^{-1} \sigma^2 \Sigma^{-1}(X_1 - X_2) \\
&\leq -\lambda_{\min}(\sigma_\Sigma^2) |X_1 - X_2|_\Sigma^2 = -2\alpha_1 h_1(X_1, X_2).
\end{aligned}$$

Then

$$\begin{aligned}
n \sum_{J \in \mathcal{J}} |g^J(x_1) - g^J(x_2)| |J|_\Sigma^2 &\leq L_1(\Lambda/\lambda_{\min}(\Sigma)) |X_1 - X_2|_\Sigma \sqrt{\lambda_{\max}(\Sigma)} \\
&\leq \alpha_1 |X_1 - X_2|_\Sigma^2 = \alpha_1 h_1(X_1, X_2),
\end{aligned}$$

if

$$|X_1 - X_2|_\Sigma \geq dL_1(\bar{\Lambda}/\alpha_1) \sqrt{\rho(\Sigma)/\lambda_{\min}(\Sigma)} =: dK_{3.5}, \quad (3.42)$$

with $K_{3.5} \in \mathcal{K}^{(3.1)}$. From this, it follows that

$$\tilde{\mathcal{A}}_n^\delta h^{(\xi)}(X_1, X_2) \leq -\alpha_1 h_1(X_1, X_2) \leq -\frac{1}{2}\alpha_1 (h_1(X_1, X_2) + d^2 K_{3.5}^2),$$

for $\delta \leq \delta_1/3$ and for $|X_1 - X_2|_\Sigma \geq dK_{3.5}$. Taking $\xi = d^2 K_{3.5}^2$ proves the lemma. \square

We now convert the drift inequality into a bound on the distribution of the coupling time

$$\tau_C := \inf\{t \geq 0: X_{n,1}^\delta(t) = X_{n,2}^\delta(t)\},$$

for arbitrary values of $(X_{n,1}^\delta(0), X_{n,2}^\delta(0))$. Our broad strategy is as follows. If $\max\{|X_{n,1}^\delta(0) - nc|_\Sigma, |X_{n,2}^\delta(0) - nc|_\Sigma\} > 3n\delta/8$, we run both processes *independently* for a fixed time interval t_1 , chosen in such a way that we have $\max\{|X_{n,1}^\delta(t_1) - nc|_\Sigma, |X_{n,2}^\delta(t_1) - nc|_\Sigma\} \leq 3n\delta/8$, with probability at least $1/16$. If not, we continue to repeat the procedure, over intervals of length t_1 , until both $X_{n,1}^\delta$ and $X_{n,2}^\delta$ are within $3n\delta/8$ of nc in the $|\cdot|_\Sigma$ -norm. We then couple the processes $X_{n,1}^\delta$ and $X_{n,2}^\delta$ as for Lemma 3.5, and run them until the minimum $(\tau_3 \wedge \tau_0)$ of the time τ_3 , at which $|X_{n,1}^\delta(t) - X_{n,2}^\delta(t)|_\Sigma$ first falls below

the value $dK_{3.5}$, and the time τ_0 , at which first $\max\{|X_{n,1}^\delta(t) - nc|_\Sigma, |X_{n,2}^\delta(t) - nc|_\Sigma\} > n\delta/2$. We call these two stages together a ‘drift phase’.

If τ_0 is the first to occur, we begin another drift phase. If not, we enter a ‘trial phase’, of length $t_3 = 1/\alpha_1$. By Theorem 3.1 and Assumption S2, and because $|X_{n,1}^\delta(t) - X_{n,2}^\delta(t)|_\Sigma \leq dK_{3.5}$ implies that $|X_{n,1}^\delta(t) - X_{n,2}^\delta(t)|_1 \leq d^{3/2}K_{3.5}\sqrt{\lambda_{\max}(\Sigma)}$, we have

$$d_{\text{TV}}(\mathcal{L}(X_{n,1}^\delta(\tau_3 + t_3)), \mathcal{L}(X_{n,2}^\delta(\tau_3 + t_3))) \leq d^{3/2}C_*n^{-1/2} \left(\frac{d\bar{\Lambda}}{g_*} \right)^{1/4}, \quad (3.43)$$

for $n \geq n_{(3.1)}$, where

$$C_* := K_{3.5}\sqrt{\lambda_{\max}(\Sigma)} \max_{1 \leq j \leq d} \left\{ K_{3.1}^j \max \left(1, \frac{\sqrt{\alpha_1}}{(dg^{(j)}\bar{\Lambda})^{1/4}} \right) \right\} \in \mathcal{K}^{(3.1)}. \quad (3.44)$$

Hence the two processes can be coupled in such a way that $X_{n,1}^\delta(\tau_3 + t_3) = X_{n,2}^\delta(\tau_3 + t_3)$, except on an event of probability at most $d^{3/2}C_*n^{-1/2}(\Lambda/g_*)^{1/4}$, and this is the coupling that we use. On the event that the values of the two processes are equal at time $\tau_3 + t_3$, the coupling is said to have been successful. If not, a new drift phase begins, (or another trial phase, if $|X_{n,1}^\delta(\tau_3 + t_3) - X_{n,2}^\delta(\tau_3 + t_3)|_\Sigma \leq dK_{3.5}$ and $\max\{|X_{n,1}^\delta(\tau_3 + t_3) - nc|_\Sigma, |X_{n,2}^\delta(\tau_3 + t_3) - nc|_\Sigma\} \leq n\delta/2$). This sequence of steps is repeated until coupling is achieved.

Theorem 3.6. *Let X_n be a sequence of Markov population processes, whose transition rates are given in (1.4), and such that Assumptions G0, G1 and S2–S4 are satisfied. Let $\delta_1 := \min\{3K_{3.5}/L_1, \delta_0/\sqrt{\lambda_{\max}(\Sigma)}\}$, and let α_1 be as in (2.3). Then, for any $\delta \leq \delta_1/3$, there is a constant $n_{3.6}$ in $\mathcal{K}^{(3.1)}$ such that, whatever the values of $X_1, X_2 \in \tilde{B}_{n,\delta}(c) := \{X \in \mathbb{Z}^d : |X - nc|_\Sigma \leq n\delta\}$ and $t \geq 0$,*

$$d_{\text{TV}}(\mathcal{L}(X_n^\delta(t) | X_n^\delta(0) = X_1), \mathcal{L}(X_n^\delta(t) | X_n^\delta(0) = X_2)) \leq 9(2n)^{1/16}e^{-\alpha_2 t},$$

for all $n \geq \max\{d^4, n_{3.6}, \psi^{-1}(\delta/2)\}$, where $\alpha_2 := \alpha_1/128$. The quantity $n_{3.6}$ is defined in (3.57).

Remark 3.7. *We prove a simple bound in Theorem 3.6 that only involves the quantities in $\mathcal{K}^{(3.1)}$ through α_1 . For this, the radius $n\delta$ of the state space of X_n^δ needs to be restricted, and this explains the upper bound on δ involving $K_{3.5}$, imposed through that on δ_1 . Since we think of δ being chosen to be small, such upper bounds are of no great importance.*

Proof. Recalling the definition (1.11) of $\tilde{B}_{n,\delta}(c)$, we begin by writing

$$\begin{aligned} B_0 &:= \tilde{B}_{n,\delta}(c) \times \tilde{B}_{n,\delta}(c); \\ B_1 &:= \{(X_1, X_2) \in B_0 : \max_{i=1,2} |X_i - nc|_\Sigma \leq n\delta/2\}; \\ B_2 &:= \{(X_1, X_2) \in B_0 : \max_{i=1,2} |X_i - nc|_\Sigma \leq 3n\delta/8\}; \\ B_3 &:= \{(X_1, X_2) \in B_1 : |X_1 - X_2|_\Sigma \leq dK_{3.5}\}. \end{aligned}$$

Clearly, $B_3 \subset B_1 \subset B_0$ and $B_2 \subset B_1$. Let

$$\begin{aligned} \tau_0 &:= \inf\{t \geq 0 : (X_{n,1}^\delta(t), X_{n,2}^\delta(t)) \notin B_1\}; \\ \tau_2 &:= \inf\{t \geq 0 : (X_{n,1}^\delta(t), X_{n,2}^\delta(t)) \in B_2\}; \\ \tau_3 &:= \inf\{t \geq 0 : (X_{n,1}^\delta(t), X_{n,2}^\delta(t)) \in B_3\}. \end{aligned}$$

Then, for any $s, \beta > 0$, and for $0 \leq i \leq 4$, we define

$$\varphi_i(\beta, s) := \max_{(X_1, X_2) \in B_i} \mathbb{E}_{X_1, X_2} \{e^{\beta(\tau_C \wedge s)}\}, \quad (3.45)$$

where the coupling time τ_C is defined for copies of the processes $X_{n,1}^\delta$ and $X_{n,2}^\delta$ specified using drift and trial phases as above, and where \mathbb{E}_{X_1, X_2} and \mathbb{P}_{X_1, X_2} refer to the distribution conditional on $(X_{n,1}^\delta(0), X_{n,2}^\delta(0)) = (X_1, X_2)$. We shall establish that, for n large enough,

$$\mathbb{P}_{X_1, X_2}[\tau_C > t] \leq 9(2n)^{1/16} e^{-\alpha_1 t/128}.$$

Fix t'_1 such that $e^{t'_1} = 128$, and write $t_1 := \alpha_1^{-1} t'_1$. Then, for any $(X_1, X_2) \in B_0$, it follows from Lemma 2.4 that

$$\mathbb{P}_{X_1, X_2}[\inf\{t > 0 : |X_{n,1}^\delta(t)|_\Sigma \leq n\delta/4\} > t_1] \leq 1/2,$$

if also $n\delta \geq 8J_{\max}^\Sigma$; this is true, from (3.2), for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta/2)\}$. But now, taking $\eta = 3\delta/8$ and $|X_0 - nc|_\Sigma \leq \delta/4$ in Lemma 2.5, it follows that

$$\mathbb{P}_{X_1, X_2}[|X_{n,1}^\delta(t_1)|_\Sigma \leq 3\delta/8] \geq 1/4,$$

provided that $n \geq \max\{n_{2.2}, (t'_1 K_{2.5}(\bar{\Lambda}/\alpha_1))^{1/2}\}$, and that $2n^{-4} \leq 1/2$. Hence, for any choice of $(X_1, X_2) \notin B_2$ and for $n \geq \max\{n_1, \psi^{-1}(\delta/2)\}$, where

$$n_1 := \max\{d^4, n_{(3.1)}\} \in \mathcal{K}^{(3.1)},$$

it follows that $\mathbb{P}_{X_1, X_2}[(X_{n,1}^\delta(t_1), X_{n,2}^\delta(t_1)) \in B_2] \geq 1/16$, if $X_{n,1}^\delta$ and $X_{n,2}^\delta$ are run independently over the interval $[0, t_1]$. Thus, defining

$$\varphi(\beta, s) := \max_{(X_1, X_2) \in B_0} \mathbb{E}_{X_1, X_2} \{e^{\beta(\tau_2 \wedge s)}\},$$

the Markov property yields

$$\varphi(\beta, s) \leq e^{\beta t_1} \{1 + 15\varphi(\beta, s)\}/16.$$

Choosing $u_0 = 1/32$, so that, in particular, $15e^{u_0 t_1}/16 = 15(128)^{u_0}/16 < 31/32$, and then $\beta = u_0 \alpha_1$, it follows for any $s > 0$ that

$$\varphi(\beta, s) \leq 2(128)^{u_0} \leq 31/15.$$

Considering the possibilities if $\tau_C \leq \tau_2$ or if $\tau_C > \tau_2$, it now follows by the strong Markov property that

$$\varphi_0(u_0 \alpha_1, s) \leq 31\varphi_2(u_0 \alpha_1, s)/15. \quad (3.46)$$

We next consider what happens if the process starts in B_2 . If $dK_{3.5} \geq 3n\delta/4$, then $B_2 \subset B_3$, and so $\varphi_2(u_0 \alpha_1, s) \leq \varphi_3(u_0 \alpha_1, s)$. If not, note that

$$\begin{aligned} & \mathbb{E}_{X_1, X_2} \{e^{\beta(\tau_C \wedge s)}\} \\ & \leq \mathbb{E}_{X_1, X_2} \{e^{\beta(\tau_C \wedge s)} \mathbb{1}\{\tau_3 \leq (\tau_0 \wedge s)\}\} + \mathbb{E}_{X_1, X_2} \{e^{\beta(\tau_C \wedge s)} \mathbb{1}\{\tau_0 \leq (\tau_3 \wedge s)\}\} \\ & \quad + \mathbb{E}_{X_1, X_2} \{e^{\beta(\tau_C \wedge s)} \mathbb{1}\{(\tau_0 \wedge \tau_3) > s\}\} \\ & \leq \mathbb{E}_{X_1, X_2} \{e^{\beta\tau_3} \mathbb{1}\{\tau_3 \leq (\tau_0 \wedge s)\}\} \varphi_3(\beta, s) \\ & \quad + e^{\beta s} \{\mathbb{P}_{X_1, X_2}[\tau_0 \leq (\tau_3 \wedge s)] + \mathbb{P}_{X_1, X_2}[(\tau_0 \wedge \tau_3) > s]\} \varphi_0(\beta, s), \end{aligned} \quad (3.47)$$

with the last inequality following from the strong Markov property. Now, in view of Lemma 3.5, if we define

$$M_1(t) := e^{\alpha_1 t/2} h_2(X_{n,1}^\delta(t), X_{n,2}^\delta(t)), \quad (3.48)$$

then $M_1(t \wedge \tau_3 \wedge \tau_0)$ is a supermartingale. This implies that

$$d^2 K_{3.5}^2 \mathbb{E}_{X_1, X_2} \{e^{\alpha_1 \tau_3/2} \mathbb{1}\{\tau_3 \leq (\tau_0 \wedge s)\}\} \leq h_2(X_1, X_2), \quad (3.49)$$

and also that

$$d^2 K_{3.5}^2 e^{\alpha_1 s/2} \mathbb{P}_{X_1, X_2}[(\tau_3 \wedge \tau_0) > s] \leq h_2(X_1, X_2). \quad (3.50)$$

Thence, by Jensen's inequality and (3.49), for any $0 \leq u \leq 1$, we also have

$$\mathbb{E}_{X_1, X_2} \{e^{u\alpha_1 \tau_3/2} \mathbb{1}\{\tau_3 \leq (\tau_0 \wedge s)\}\} \leq \{h_2(X_1, X_2)/d^2 K_{3.5}^2\}^u. \quad (3.51)$$

Finally, from Lemma 2.5 with $\eta = \delta/2$, for $(X_1, X_2) \in B_2$ and for θ_1 as in Lemma 2.2,

$$\mathbb{P}_{X_1, X_2}[\tau_0 \leq (\tau_3 \wedge s)] \leq \mathbb{P}_{X_1, X_2}[\tau_0 \leq s] \leq 2 \exp\{-7n\theta_1 \delta^2/64\}, \quad (3.52)$$

if $s \leq n/\alpha_1$ and $n \geq \max\{n_2, \psi^{-1}(\delta/2)\}$, where

$$n_2 := \max \left\{ n_{(3.1)}, \frac{K_{2.5}\bar{\Lambda}}{\alpha_1} \right\} \in \mathcal{K}^{(3.1)};$$

this follows because $n_{2.2} \leq n_{(3.1)}$, and because, for n and s chosen in this way, $\exp\{n^{-1}\theta_1(3n\delta/8)^2\} \geq n^9 \geq nK_{2.5}\Lambda s$, because $\delta \geq 2\psi(n)$ and $n \geq d^4$.

Hence, taking $\beta = u_0\alpha_1/2$, with $u_0 = 1/32$ as above, substituting (3.50), (3.51) and (3.52) into (3.47), and then using (3.46), we have

$$\begin{aligned} \varphi_2(u_0\alpha_1/2, s) &\leq (2n\delta/dK_{3.5})^{2u_0}\varphi_3(u_0\alpha_1/2, s) + P(u_0, s, n)\varphi_0(u_0\alpha_1/2, s) \\ &\leq (2n\delta/dK_{3.5})^{2u_0}\varphi_3(u_0\alpha_1/2, s) + 31P(u_0, s, n)\varphi_2(u_0\alpha_1/2, s)/15, \end{aligned}$$

where

$$\begin{aligned} P(u, s, n) &:= e^{u\alpha_1 s/2} \{ \mathbb{P}_{X_1, X_2}[\tau_0 \leq (\tau_3 \wedge s)] + \mathbb{P}_{X_1, X_2}[(\tau_3 \wedge \tau_0) > s] \} \\ &\leq e^{u\alpha_1 s/2} \{ 2 \exp\{-7n^{3/4}d\theta_1\delta^2/64\} + (2n\delta/dK_{3.5})^{2u_0}e^{-\alpha_1 s/2} \}, \end{aligned}$$

recalling $n \geq d^4$ for the final inequality. It thus follows that $P(u_0, s_n, n) \leq 15/62$ for all $n \geq \max\{n_2, d^4, \psi^{-1}(\delta/2)\}$ such that

$$\left(\frac{2n\delta}{dK_{3.5}} \right)^{2u_0} e^{-\alpha_1(1-u_0)s_n/2} \leq \frac{7}{62} \quad \text{and} \quad e^{-(7n^{3/4}d\theta_1\delta^2/64 - u_0\alpha_1 s_n/2)} \leq \frac{2}{31}. \quad (3.53)$$

Picking $s = s_n := 64 \log n / \alpha_1$, and recalling that $u_0 = 1/32$ and $2\psi(n) \leq \delta \leq \delta_1/3 \leq K_{3.5}/L_1$, it is enough that $n \geq n_3$, where

$$n_3 := \max \left\{ \left[\left(\frac{2}{L_1} \right)^{1/16} \left(\frac{62}{7} \right) \right]^{1/30}, \left(\frac{31}{2} \right)^{1/6} \right\} \in \mathcal{K}^{(3.1)}.$$

Hence, for $n \geq \max\{n_2, n_3, \psi^{-1}(\delta/2)\}$,

$$\varphi_2(u_0\alpha_1/2, s_n) \leq 2(2n\delta/dK_{3.5})^{2u_0}\varphi_3(u_0\alpha_1/2, s_n). \quad (3.54)$$

For $(X_1, X_2) \in B_3$, we take $t_3 = 1/\alpha_1$, and use (3.43) to conclude that

$$\varphi_3(u\alpha_1/2, s_n) \leq e^{u/2} \{ 1 + d^{3/2}C_*n^{-1/2}(\Lambda/g_*)^{1/4} \varphi_0(u\alpha_1/2, s_n) \}, \quad (3.55)$$

for C_* as defined in (3.44), if $n \geq n_{(3.1)}$. From (3.46) and (3.54), we have

$$\varphi_0(u_0\alpha_1/2, s_n) \leq 62(2n\delta/dK_{3.5})^{2u_0}\varphi_3(u_0\alpha_1/2, s_n)/15, \quad (3.56)$$

for all $n \geq \max\{d^4, n_2, n_3, \psi^{-1}(\delta/2)\}$. Taking $u = u_0 = 1/32$ in (3.55) and using (3.56), the coefficient of $\varphi_3(u_0\alpha_1/2, s_n)$ on the right hand side of (3.55) is at most

$$e^{1/64}d^{3/2}C_*n^{-1/2}(\Lambda/g_*)^{1/4}62(2n\delta/dK_{3.5})^{1/16}/15 \leq 1/2$$

if, using $n \geq d^4$ and $\delta_1/3 \leq K_{3.5}/L_1$,

$$n \geq n_4 := e \left(\frac{2}{L_1} \right)^4 \left(\frac{124C_*}{15} \right)^{64} \left(\frac{\bar{\Lambda}}{g_*} \right)^{16} \in \mathcal{K}^{(3.1)}.$$

Hence, from (3.55) and (3.56), for $n \geq \max\{\max_{2 \leq l \leq 4} n_l, \psi^{-1}(\delta/2)\}$, we have

$$\varphi_3(u_0\alpha_1/2, s_n) \leq 2e^{1/64}.$$

Combining this with (3.56) and the definition (3.45) of φ_0 , it follows that, for all $(X_1, X_2) \in B_0$, we have

$$\mathbb{P}_{X_1, X_2}[\tau_C > t] \leq 9(2n)^{2u_0}e^{-u_0\alpha_1 t/2}, \quad 0 \leq t \leq s_n,$$

for $n \geq \max\{n_{(3.1)}, \max_{1 \leq l \leq 4} n_l, \psi^{-1}(\delta/2)\}$. In particular,

$$\mathbb{P}_{X_1, X_2}[\tau_C > s_n] \leq 9(2n)^{2u_0}e^{-u_0\alpha_1 s_n/2}.$$

However, by the strong Markov property, for $t > s_n$,

$$\mathbb{P}_{X_1, X_2}[\tau_C > t] \leq \mathbb{P}_{X_1, X_2}[\tau_C > s_n] \max_{(X_1, X_2) \in B_0} \mathbb{P}_{X_1, X_2}[\tau_C > t - s_n].$$

Arguing inductively, it follows that, for $r \in \mathbb{Z}_+$ and $0 \leq v < s_n$,

$$\mathbb{P}_{X_1, X_2}[\tau_C > rs_n + v] \leq (9(2n)^{2u_0}e^{-u_0\alpha_1 s_n/2})^r 9(2n)^{2u_0}e^{-u_0\alpha_1 v/2}.$$

Take n so large that $9(2n)^{2u_0}e^{-u_0\alpha_1 s_n/4} \leq 1$; this is true, for $u_0 = 1/32$ and $s_n = 64 \log n / \alpha_1$, if $n \geq n_5 := 9^4 2^{1/4}$. Then, for all $(X_1, X_2) \in B_0$, we have

$$\mathbb{P}_{X_1, X_2}[\tau_C > rs_n + v] \leq 9(2n)^{2u_0}e^{-u_0\alpha_1(rs_n+v)/4}.$$

Since $u_0 = 1/32$, the inequality in the theorem is thus proved for $n \geq \max\{n_{3.6}, \psi^{-1}(\delta/2)\}$, where

$$n_{3.6} := \max_{1 \leq l \leq 5} n_l \in \mathcal{K}^{(3.1)}, \quad (3.57)$$

with $\alpha_1/128$ for α_2 . □

4 Stein's method based on X_n^δ

4.1 Bounding the solutions of the Stein equation

We now use the results of the previous section to bound the first and second differences of the solutions $h_B := h_{B,n}^\delta$ of the Stein equation corresponding to the generator \mathcal{A}_n^δ defined in (1.12), for the special Markov population processes satisfying Assumptions G0, G1 and S2–S4. We recall from the introduction the definitions

$$\|\Delta f(X)\|_\infty := \max_{1 \leq j \leq d} |\Delta_j f(X)|; \quad \|\Delta^2 f(X)\|_\infty := \max_{1 \leq j, k \leq d} |\Delta_{jk} f(X)|, \quad (4.1)$$

where Δ_j and Δ_{jk} , as defined in (1.15), denote the components of the first and second difference operators Δ and Δ^2 , respectively.

Theorem 4.1. *Let X_n be a Markov population process whose transition rates are given in (1.4), satisfying Assumptions G0, G1 and S2–S4 for some $\delta_0 > 0$. Let $\delta_1 := \min\{3K_{3.5}/L_1, \delta_0/\sqrt{\lambda_{\max}(\Sigma)}\}$ be as in Theorem 3.6. Then there are constants $\kappa_0, \kappa_1, \kappa_2 \in \mathcal{K}^{(3.1)}$ such that, for any $B \subset \mathbb{Z}^d$ and any $\delta \leq \delta_1/3$, the solution $h_B := h_{B,n}^\delta$ of the Stein equation*

$$\mathcal{A}_n^\delta h_B(X) = \mathbb{1}_B(X) - \Pi_n^\delta\{B\}$$

satisfies

$$\begin{aligned} |h_B(X)| &\leq \alpha_1^{-1} \kappa_0 \log n; \quad \|\Delta h_B(X)\|_\infty \leq \alpha_1^{-1} \kappa_1 d^{1/4} n^{-1/2} \log n; \\ \|\Delta^2 h_B(X)\|_\infty &\leq \alpha_1^{-1} \kappa_2 d^{1/2} n^{-1} \log n, \end{aligned}$$

for all $|X - nc|_\Sigma \leq n\delta/4$ and $n \geq \max\{n_{3.6}, \psi^{-1}(\delta/2)\}$. The constants $\kappa_0, \kappa_1, \kappa_2$ are given in (4.3), (4.5) and (4.6), respectively.

Proof. The argument starts from the explicit representation of h_B given in (1.16), which immediately yields

$$\begin{aligned} h_B(X) &= - \int_0^\infty (\mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = X] - \Pi_n^\delta\{B\}) dt \\ &= - \sum_{Y \in \tilde{B}_{n,\delta}(c)} \int_0^\infty (\mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = X] \\ &\quad - \mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = Y]) \Pi_n^\delta(Y) dt. \end{aligned} \quad (4.2)$$

Because $\delta \leq \delta_1/3$ and $9 \cdot 2^{1/16} \leq 10$, we can use Theorem 3.6 for $n \geq \max\{n_{3.6}, \psi^{-1}(\delta/2)\}$ to give

$$\begin{aligned} |h_B(X)| &\leq 2\alpha_2^{-1} \log n + \int_{2\alpha_2^{-1} \log n}^{\infty} 10n^{1/16} e^{-\alpha_2 t} dt \\ &\leq 2\alpha_2^{-1} \log n + 10\alpha_2^{-1} n^{-1}, \end{aligned}$$

proving the bound on $|h_B(X)|$, with

$$\kappa_0 := 1536. \quad (4.3)$$

Next, from (4.2), we have

$$\begin{aligned} h_B(X - e^{(j)}) - h_B(X) &= \int_0^{\infty} (\mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = X] - \mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = X - e^{(j)}]) dt. \end{aligned} \quad (4.4)$$

Taking $f(X) := \mathbb{1}_B(X)$, Theorem 3.1 implies that

$$\begin{aligned} &\int_0^{2\alpha_2^{-1} \log n} |\mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = X] - \mathbb{P}[X_n^\delta(t) \in B \mid X_n^\delta(0) = X - e^{(j)}]| dt \\ &\leq K_{3.1}^j \alpha_2^{-1} n^{-1/2} \{2 \log n + 2\alpha_2 (\Lambda g^{(j)})^{-1/2}\} \left(\frac{\Lambda}{g^{(j)}} \right)^{1/4}, \end{aligned}$$

if $|X - nc|_\Sigma \leq n\delta/4$ and $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta/2)\}$; in performing the integration, the range is split at $t = (\Lambda g^{(j)})^{-1/2}$. The remainder of the integral is bounded by $10\alpha_2^{-1} n^{-1}$, as above, if also $n \geq \max\{n_{3.6}, \psi^{-1}(\delta/2)\}$, since $n_{3.6} \geq n_{(3.1)}$, completing the bound on $|h(X - e^{(j)}) - h(X)|$, and thence on $\|\Delta h(X)\|_\infty$, with

$$\kappa_1 := 128 \left\{ 10 + K_*^{(1)} \left(2 + \frac{2\alpha_2}{\sqrt{\Lambda g_*}} \right) \right\} \left(\frac{\bar{\Lambda}}{g_*} \right)^{1/4} \in \mathcal{K}^{(3.1)}, \quad (4.5)$$

where $K_*^{(1)} := \max_{1 \leq j \leq d} K_{3.1}^j$.

For the second differences, the argument is entirely similar, using Theorem 3.3 for the bulk of the estimate, and bounding the integrand by 2 for $0 \leq t \leq n^{-1}(\Lambda g_{ij}^+)^{-1/2}$. This gives the bound on $\|\Delta^2 h(X)\|_\infty$, with

$$\kappa_2 := 128 \left\{ 10 + \frac{2\alpha_2}{\sqrt{\Lambda g_*}} + K_*^{(2)} \left(\frac{\bar{\Lambda}}{g_*} \right)^{1/2} \left(2 + \frac{\alpha_2}{\sqrt{\Lambda g_*}} \right) \right\} \in \mathcal{K}^{(3.1)}, \quad (4.6)$$

where $K_*^{(2)} := \max_{1 \leq i, j \leq d} K_{3.3}^{ji}$. □

Remark 4.2. As a consequence of Remark 3.4, the dimension d appearing in the factors $d^{1/4}$ and $d^{1/2}$ in the bounds on $\|\Delta h_B(X)\|_\infty$ and $\|\Delta^2 h_B(X)\|_\infty$, respectively, can be replaced by $c(A)g^*/\bar{\Lambda}$. Thus, for instance, if we have $A = -\lambda I$, for some $\lambda > 0$, as in Example 6.3, then d can be replaced by the dimension independent constant $g^*/\bar{\Lambda}$ in these factors.

4.2 Reducing the generator

In this section, we show that the generator \mathcal{A}_n^δ can be replaced for our purposes by the simpler operator \mathcal{A}_n , defined in (1.1). As a first step in the reduction, we use two technical lemmas to bound the expectation of the Newton remainder

$$e_2(W, J, h) := h(W + J) - h(W) - \Delta h(W)^T J - \frac{1}{2} J^T \Delta^2 h(W) J, \quad (4.7)$$

for W a random vector. For $\eta > 0$ and $f: \mathbb{Z}^d \rightarrow \mathbb{R}$, we use the notation

$$\|f\|_{n\eta, \infty}^\Sigma := \max_{|X - nc|_\Sigma \leq n\eta} |f(X)|, \quad (4.8)$$

analogous to that of (1.17), but using $|\cdot|_\Sigma$ -balls, with nc implicit. Similarly, we write $\|\Delta h\|_{n\eta, \infty}^\Sigma$ and $\|\Delta^2 h\|_{n\eta, \infty}^\Sigma$ for $\|f\|_{n\eta, \infty}^\Sigma$, when $f(X) = \|\Delta h(X)\|_\infty$ and $f(X) = \|\Delta^2 h(X)\|_\infty$, respectively, so that the conclusion of Theorem 4.1 can be expressed as

$$\begin{aligned} \|h_B\|_{n\delta/4, \infty}^\Sigma &\leq \alpha_1^{-1} \kappa_0 \log n; & \|\Delta h_B\|_{n\delta/4, \infty}^\Sigma &\leq \alpha_1^{-1} \kappa_1 d^{1/4} n^{-1/2} \log n; \\ \|\Delta^2 h_B\|_{n\delta/4, \infty}^\Sigma &\leq \alpha_1^{-1} \kappa_2 d^{1/2} n^{-1} \log n, \end{aligned} \quad (4.9)$$

for $n \geq \max\{n_{3.6}, \psi^{-1}(\delta/2)\}$ and $\delta \leq \delta_1$.

Our control over the differences of the functions h_B is only such that we can bound their $\|\cdot\|_{n\eta, \infty}^\Sigma$ norms for suitable η , so these are the quantities that we need in our estimates. For instance, if W is a random vector in \mathbb{Z}^d such that $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon$, we immediately have the bound

$$|\mathbb{E}\{f(W + e^{(j)}) - f(W)\}| \leq 2\varepsilon \|f\|_\infty.$$

Because we often only have control within certain (large) balls, we are led instead to bounding a truncated quantity

$$|\mathbb{E}\{(f(W + e^{(j)}) - f(W))I[|W - nc|_\Sigma \leq n\eta_1]\}|$$

in terms of $\|f\|_{n\eta_2, \infty}^\Sigma$, for suitable choices of η_1 and η_2 . The following lemma, proved in Appendix 7.1, provides what we need.

Lemma 4.3. *Suppose that W is a random vector in \mathbb{Z}^d such that*

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon_1, \quad 1 \leq j \leq d; \quad \mathbb{P}[|W - nc|_{\Sigma} > n\delta/4] \leq \varepsilon_2.$$

Then, for $f: \mathbb{Z}^d \rightarrow \mathbb{R}$ and $1 \leq j \leq d$, and for any $U, X \in \mathbb{Z}^d$ such that $\max\{|X|_{\Sigma}, |X + U|_{\Sigma}\} \leq n\delta/6$ and $n\delta \geq 12|U|_{\Sigma}$,

$$\begin{aligned} & |\mathbb{E}\{(f(W + X + U) - f(W + X))I[|W - nc|_{\Sigma} \leq n\delta/3]\}| \\ & \leq (\varepsilon_1|U|_1 + \varepsilon_2)\|f\|_{n\delta/2, \infty}^{\Sigma}. \end{aligned}$$

We use Lemma 4.3 to bound the Newton remainder defined in (4.7). Instead of bounding $e_2(W, J, h)$ directly, we bound a perturbation of it,

$$E_2(W, J, h) := e_2(W, J, h) + \frac{1}{2} \sum_{j=1}^d J_j \Delta_{jj} h(W), \quad (4.10)$$

which can be represented as a sum of third differences of h . The result, proved in Appendix 7.2, is as follows.

Lemma 4.4. *If W is a random vector in \mathbb{Z}^d such that*

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon_1, \quad 1 \leq j \leq d; \quad \mathbb{P}[|W - nc|_{\Sigma} \geq n\delta/4] \leq \varepsilon_2,$$

then, for any $J \in \mathbb{Z}^d$ such that $|J|_{\Sigma} \leq n\delta/12$,

$$\begin{aligned} (i) \quad & \left| \mathbb{E}\left\{E_2(W, J, h)I[|W - nc|_{\Sigma} \leq n\delta/3]\right\} \right| \\ & \leq \|\Delta^2 h\|_{n\delta/2, \infty}^{\Sigma} (C_{4.4}^{(1)}(J)\varepsilon_1 + C_{4.4}^{(2)}(J)\varepsilon_2), \end{aligned}$$

where

$$C_{4.4}^{(1)}(J) := \frac{1}{6}|J|_1(|J|_1 + 1)(|J|_1 + 2); \quad C_{4.4}^{(2)}(J) := \frac{1}{2}|J|_1(|J|_1 + 1). \quad (4.11)$$

If the conditions are replaced by $\mathbb{P}[|W - nc| \geq n\delta/4] \leq \varepsilon_2^E$ and $|J| \leq n\delta/12$, then

$$\begin{aligned} (ii) \quad & \left| \mathbb{E}\left\{E_2(W, J, h)I[|W - nc| \leq n\delta/3]\right\} \right| \\ & \leq \|\Delta^2 h\|_{n\delta/2, \infty} (C_{4.4}^{(1)}(J)\varepsilon_1 + C_{4.4}^{(2)}(J)\varepsilon_2^E), \end{aligned}$$

where $\|\Delta^2 h\|_{n\delta/2, \infty}$ is as defined in (1.17).

Remark 4.5. *Note that, because $0 \neq J \in \mathbb{Z}^d$, we have*

$$C_{4.4}^{(1)}(J) \leq |J|_1^3 \leq d^{3/2}|J|^3; \quad C_{4.4}^{(2)}(J) \leq 2d|J|^2.$$

Lemma 4.4 allows us to prove the following reduction theorem, useful for approximating the generator of any Markov population process satisfying our general assumptions.

Theorem 4.6. *Suppose that $(g^J, J \in \mathcal{J})$, c , A , σ^2 , γ , δ_0 , \mathcal{A}_n^δ and \mathcal{A}_n are as in Sections 1 and 2.1, and that Assumptions G0–G4 are satisfied. Suppose that W is a random vector in \mathbb{Z}^d , such that, for some $\varepsilon, V > 0$,*

$$\begin{aligned} \text{(i)} \quad & \mathbb{E}|W - nc|_\Sigma^2 \leq dVn; \\ \text{(ii)} \quad & d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon, \text{ for each } 1 \leq j \leq d. \end{aligned} \tag{4.12}$$

Then, for any $\delta \leq \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$ and for any $0 < \delta' \leq \delta/2$, and for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta/2)\}$,

$$\begin{aligned} & |\mathbb{E}\{(\mathcal{A}_n^\delta h(W) - \mathcal{A}_n h(W))I[|W - nc|_\Sigma \leq n\delta'/3]\}| \\ & \leq d^{5/2}\bar{\Lambda} \left(\frac{1}{2}L_2\lambda_{\max}(\Sigma)V\|\Delta h\|_{n\delta/4,\infty}^\Sigma \right. \\ & \quad \left. + n\|\Delta^2 h\|_{n\delta/4,\infty}^\Sigma \{L_1\sqrt{V}n^{-1/2} + d^{1/2}(\bar{\gamma}/\bar{\Lambda})\varepsilon + 32d^{1/2}V/\{n(\delta')^2\}\} \right), \end{aligned}$$

where $\|\Delta h\|_{n\eta,\infty}^\Sigma$ and $\|\Delta^2 h\|_{n\eta,\infty}^\Sigma$ are bounded in (4.9).

Proof. Consider

$$\begin{aligned} \mathcal{A}_n^\delta h(X) &:= n \sum_{J \in \mathcal{J}} g_\delta^J(X/n) \{h(X+J) - h(X)\} \\ &= n \sum_{J \in \mathcal{J}} \{g^J(c) + Dg^J(c)^T n^{-1}(X - nc) + e_1(X, J, g_\delta^J)\} \\ & \quad \times \{\Delta h(X)^T J + \tfrac{1}{2}J^T \Delta^2 h(X)J + e_2(X, J, h)\}, \end{aligned}$$

where

$$e_1(X, J, g_\delta^J) := g_\delta^J(X/n) - g^J(c) - n^{-1}Dg^J(c)^T(X - nc),$$

and e_2 is as in (4.7). Observing that $\sum_{J \in \mathcal{J}} g^J(c)\Delta h(X)^T J = \Delta h(X)^T F(c) = 0$, because $F(c) = 0$, and that

$$\begin{aligned} \sum_{J \in \mathcal{J}} g^J(c)J^T \Delta^2 h(X)J &= \text{Tr}\{\sigma^2 \Delta^2 h(X)\}; \\ \sum_{J \in \mathcal{J}} Dg^J(c)^T n^{-1}(X - nc)\Delta h(X)^T J &= n^{-1}\text{Tr}\{A(X - nc)\Delta h(X)^T\}, \end{aligned}$$

it follows, once again writing $I_n^\eta(W) := I[|W - nc|_\Sigma \leq n\eta/3]$, that

$$\begin{aligned}
& |\mathbb{E}\{(\mathcal{A}_n^\delta h(W) - \mathcal{A}_n h(W))I_n^{\delta'}(W)\}| \\
& \leq n\mathbb{E}\left\{\sum_{J \in \mathcal{J}} |e_1(W, J, g_\delta^J)| |h(W + J) - h(W)| I_n^{\delta'}(W)\right\} \\
& \quad + \mathbb{E}\left\{\sum_{J \in \mathcal{J}} |Dg^J(c)^T(W - nc)| |h(W + J) - h(W) - \Delta h(W)^T J| I_n^{\delta'}(W)\right\} \\
& \quad + n\left|\mathbb{E}\left\{\sum_{J \in \mathcal{J}} g^J(c) e_2(W, J, h) I_n^{\delta'}(W)\right\}\right|. \tag{4.13}
\end{aligned}$$

Now, from (2.2), and recalling (1.9), it follows that, for $X \in \mathcal{X}_n^\delta(J)$,

$$\begin{aligned}
|e_1(X, J, g_\delta^J)| &= |g_\delta^J(X/n) - g^J(c) - Dg^J(c)^T n^{-1}(X - nc)| \\
&\leq \frac{1}{2}n^{-2}|X - nc|^2 L_2 g^J(c), \tag{4.14}
\end{aligned}$$

provided that $\delta\sqrt{\lambda_{\max}(\Sigma)} \leq \delta_0$. Since, for $|X - nc|_\Sigma \leq n\delta'/3 \leq n\delta/6$ and $n\delta/12 > J_{\max}^\Sigma$, we have

$$|X + J - nc|_\Sigma \leq J_{\max}^\Sigma + |X - nc|_\Sigma \leq n\delta/4,$$

it follows, for such X and n , that $|h(X + J) - h(X)| \leq |J|_1 \|\Delta h\|_{n\delta/4, \infty}^\Sigma$ and that (4.14) is satisfied. Hence, since $\mathbb{E}|W - nc|_\Sigma^2 \leq dVn$, we have

$$\begin{aligned}
& n\mathbb{E}\left\{\sum_{J \in \mathcal{J}} |e_1(W, J, g_\delta^J)| |h(W + J) - h(W)| I_n^{\delta'}(W)\right\} \\
& \leq \frac{1}{2}L_2 \sum_{J \in \mathcal{J}} g^J(c) |J|_1 \|\Delta h\|_{n\delta/4, \infty}^\Sigma n^{-1} \mathbb{E}|W - nc|^2 \\
& \leq \frac{1}{2}L_2 \sum_{J \in \mathcal{J}} g^J(c) |J|_1 \|\Delta h\|_{n\delta/4, \infty}^\Sigma dV \lambda_{\max}(\Sigma) \\
& \leq \frac{1}{2}L_2 V d^{5/2} \bar{\Lambda} \lambda_{\max}(\Sigma) \|\Delta h\|_{n\delta/4, \infty}^\Sigma,
\end{aligned}$$

if $n\delta/12 > J_{\max}^\Sigma$, and this condition is satisfied for $n \geq \max\{n_{(3.1)}, \psi^{-1}(\delta/2)\}$.

Then, from (7.37) and (2.2), if $n \geq n_{(3.1)}$ and $|X - nc|_\Sigma \leq n\delta/3$,

$$\begin{aligned}
& \sum_{J \in \mathcal{J}} |Dg^J(c)^T(X - nc)| |h(X + J) - h(X) - \Delta h(X)^T J| I_n^{\delta'}(X) \\
& \leq \frac{1}{2}L_1 |X - nc| \|\Delta^2 h\|_{n\delta/4, \infty}^\Sigma \sum_{J \in \mathcal{J}} g^J(c) |J|_1 (|J|_1 + 1) \\
& \leq d^2 \bar{\Lambda} L_1 |X - nc| \|\Delta^2 h\|_{n\delta/4, \infty}^\Sigma;
\end{aligned}$$

hence, since $\mathbb{E}|W - nc|_\Sigma \leq \sqrt{dVn}$,

$$\begin{aligned} & \mathbb{E} \left\{ \sum_{J \in \mathcal{J}} |Dg^J(c)^T(W - nc)| |h(W + J) - h(W) - \Delta h(W)^T J| I_n^{\delta'}(W) \right\} \\ & \leq d^{5/2} \bar{\Lambda} L_1 \sqrt{Vn} \|\Delta^2 h\|_{n\delta/4, \infty}^\Sigma. \end{aligned}$$

Finally, from Lemma 4.4 and Chebyshev's inequality,

$$\begin{aligned} & n \left| \mathbb{E} \left\{ \sum_{J \in \mathcal{J}} g^J(c) \left(e_2(W, J, h) + \frac{1}{2} \sum_{j=1}^d J_j \Delta_{jj} h(W) \right) I_n^{\delta'}(W) \right\} \right| \\ & \leq \sum_{J \in \mathcal{J}} g^J(c) n \|\Delta^2 h\|_{n\delta/4, \infty}^\Sigma (C_{4.4}^{(1)}(J) \varepsilon + 16 C_{4.4}^{(2)}(J) dV / \{n(\delta')^2\}) \\ & \leq n \|\Delta^2 h\|_{n\delta/4, \infty}^\Sigma (d^3 \bar{\gamma} \varepsilon + 32 d^3 \bar{\Lambda} V / \{n(\delta')^2\}), \end{aligned}$$

and, for each j , $\sum_{J \in \mathcal{J}} J_j g^J(c) = 0$, because $F(c) = 0$. This completes the proof of the theorem. \square

4.3 A first total variation approximation theorem

We are now in a position to prove Theorem 4.7, which gives a measure of the error in the approximation of the distribution of a random vector W in \mathbb{Z}^d by the distribution Π_n^δ , if the process X_n^δ satisfies the special assumptions of Section 3. Using this theorem, it is shown in Section 5 that the discrete normal distribution $\mathcal{DN}_a(nc, n\Sigma)$ can be well approximated by an equilibrium distribution Π_n^δ , for a suitably chosen X_n^δ , so that Theorem 4.7 in turn implies the discrete normal approximation of Theorem 5.5. The statement of Theorem 4.7 is to some extent complicated by the presence of the indicator truncating the range of W in the main condition (iii). The truncation is necessary, because Theorem 4.1 only enables one to bound the differences of the functions h_B solving the Stein equation (1.13) in balls of radius $n\delta/4$, for any $\delta \leq \delta_0/3\sqrt{\lambda_{\max}(\Sigma)}$.

Theorem 4.7. *Suppose that $(g^J, J \in \mathcal{J})$, c , A , σ^2 , γ , δ_0 and \mathcal{A}_n^δ are as in Sections 1 and 2.1, and that Assumptions G0, G1 and S2–S4 are satisfied. Define $\tilde{\delta}_0 := \min\{K_{3.5}/L_1, \delta_0/3\sqrt{\lambda_{\max}(\Sigma)}\}$, and suppose that $\delta' \leq \tilde{\delta}_0/2$. Then, for any $V > 0$, there exists a constant $C_{4.7}(V, \delta')$, which is a function of V , δ' and the elements of $\mathcal{K}^{(3.1)}$, with the following property: if W is any random vector in \mathbb{Z}^d such that, for some $n \geq \max\{n_{3.6}, \psi^{-1}(\delta')\}$ and for*

some $\varepsilon_1, \varepsilon_{20}, \varepsilon_{21}, \varepsilon_{22} > 0$,

- (i) $\mathbb{E}|W - nc|_{\Sigma}^2 \leq dVn$;
- (ii) $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon_1$, for each $1 \leq j \leq d$;
- (iii) $|\mathbb{E}\{\mathcal{A}_n h(W) I[|W - nc|_{\Sigma} \leq n\delta'/3]\}|$
 $\leq \bar{\Lambda}(\varepsilon_{20}\|h\|_{n\tilde{\delta}_0/4, \infty}^{\Sigma} + \varepsilon_{21}n^{1/2}\|\Delta h\|_{n\tilde{\delta}_0/4, \infty}^{\Sigma} + \varepsilon_{22}n\|\Delta^2 h\|_{n\tilde{\delta}_0/4, \infty}^{\Sigma}),$

where \mathcal{A}_n is as defined in (1.1) and $\bar{\Lambda} := d^{-1}\text{Tr}(\sigma^2)$, then, for any δ such that $2\delta' \leq \delta \leq \tilde{\delta}_0$,

$$d_{\text{TV}}(\mathcal{L}(W), \Pi_n^{\delta}) \leq C_{4.7}(V, \delta')(d^3 n^{-1/2} + d^{7/2}(\bar{\gamma}/\bar{\Lambda})\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22}) \log n.$$

Proof. From (1.14), we have

$$\begin{aligned} d_{\text{TV}}(\mathcal{L}(W), \Pi_n^{\delta}) \\ \leq \sup_{B \subset \tilde{B}_{n, \delta}(c)} |\mathbb{E}\{\mathcal{A}_n^{\delta} h_B(W) I[W \in \tilde{B}_{n, \delta'/3}(c)]\}| + \mathbb{P}[W \notin \tilde{B}_{n, \delta'/3}(c)], \end{aligned}$$

where $\tilde{B}_{n, \eta}(c) := \mathbb{Z}^d \cap B_{n\eta, \Sigma}(nc)$ and $h_B := h_{B, n}^{\delta}$ is as for (1.13). The probability in the second term is at most $9dV/\{n(\delta')^2\}$, by (i) and Chebyshev's inequality. Then, for $2\delta' \leq \delta \leq \tilde{\delta}_0$ and for $n \geq \max\{n_{3.6}, \psi^{-1}(\delta')\}$, we can use (4.9) in (iii), giving

$$\begin{aligned} & |\mathbb{E}\{\mathcal{A}_n h_B(W)\} I[|W - nc|_{\Sigma} \leq n\delta'/3]| \\ & \leq \bar{\Lambda}(\varepsilon_{20}\|h_B\|_{n\tilde{\delta}_0/4, \infty}^{\Sigma} + \varepsilon_{21}n^{1/2}\|\Delta h_B\|_{n\tilde{\delta}_0/4, \infty}^{\Sigma} + \varepsilon_{22}n\|\Delta^2 h_B\|_{n\tilde{\delta}_0/4, \infty}^{\Sigma}) \\ & \leq (\bar{\Lambda}/\alpha_1)(\varepsilon_{20}\kappa_0 + \varepsilon_{21}\kappa_1 d^{1/4} + \varepsilon_{22}\kappa_2 d^{1/2}) \log n. \end{aligned}$$

Finally, from Theorem 4.6 and (4.9), for $n \geq \max\{n_{3.6}, \psi^{-1}(\delta')\}$ and $2\delta' \leq \delta \leq \tilde{\delta}_0$, we have

$$\begin{aligned} & |\mathbb{E}\{(\mathcal{A}_n^{\delta} h_B(W) - \mathcal{A}_n h_B(W)) I[W \in \tilde{B}_{n, \delta'/3}(c)]\}| \\ & \leq d^3 \log n \frac{\bar{\Lambda}}{\alpha_1} \left\{ \frac{\frac{1}{2}\kappa_1 L_2 \lambda_{\max}(\Sigma)V + \kappa_2 L_1 \sqrt{V}}{\sqrt{n}} + \frac{32\kappa_2 d^{1/2}V}{n(\delta')^2} + \kappa_2 d^{1/2}(\bar{\gamma}/\bar{\Lambda})\varepsilon_1 \right\}, \end{aligned}$$

completing the proof of the theorem. \square

5 Discrete normal approximation

In this section, we show that Theorem 4.7 can be used to establish approximation by distributions from the discrete normal family. To do so, we need first to establish properties of distributions in the family that are related to the conditions of Theorem 4.7. As in Section 2.1, we assume that $n \geq d^4$.

5.1 Properties of discrete normal distributions

We first note the following simple lemma, proved in Appendix 7.3, in which certain moments of the discrete normal random variable $W \sim \mathcal{DN}_d(nc, n\Sigma)$ are bounded by expressions similar to those of $\mathcal{N}_d(nc, n\Sigma)$.

Lemma 5.1. *For $l \in \mathbb{Z}_+$, we have*

$$(a) \quad \mathbb{E}|W - nc|_\Sigma^l \leq C(l)(nd)^{l/2},$$

whenever $n \geq 1/\lambda_{\min}(\Sigma)$, for universal constants $C(l)$. In addition, for each $1 \leq j \leq d$ and $n \geq 1$,

$$(b) \quad \mathbb{E}(W_j - nc_j)^2 \leq \frac{1}{2} + 2n\Sigma_{jj},$$

and, for $l \in \mathbb{Z}_+$ and for universal constants $C'(l)$,

$$(c) \quad \mathbb{E}\{[\Sigma^{-1}(W - nc)]_j^{2l}\} \leq n^l C'(l)(1 + (\Sigma^{-1})_{jj}^l),$$

whenever $n \geq d/\{4(\lambda_{\min}(\Sigma))^2\}$.

The next lemma, proved in Appendix 7.4, establishes an approximate integration by parts formula for multivariate discrete normal distributions. As before, we write $I_n^\eta(X) := I[|X - nc|_\Sigma \leq n\eta/3]$ for any $\eta > 0$. We say that $C \in \mathcal{K}_\Sigma$ if C is an increasing function of $\lambda_{\max}(\Sigma)$, $1/\lambda_{\min}(\Sigma)$, and $C(\delta) \in \mathcal{K}_\Sigma(\delta)$ if $C(\delta) \in \mathcal{K}_\Sigma$ for each fixed δ . By analogy with (2.19), we also define

$$\psi_\Sigma(n) := \frac{6}{n\sqrt{\lambda_{\min}(\Sigma)}}, \quad (5.1)$$

noting that then $\psi_\Sigma^{-1}(\delta) = \psi_\Sigma(\delta)$.

Lemma 5.2. *Suppose that $W \sim \mathcal{DN}_d(nc, n\Sigma)$. Then there exist constants $n_{5.2} \in \mathcal{K}$ and $C_{5.2}^{(1)}(\delta), C_{5.2}^{(2)}(\delta), C_{5.2}^{(3)}(\delta) \in \mathcal{K}_\Sigma(\delta)$, such that, for any $n \geq \max\{n_{5.2}, \psi_\Sigma(\delta)\}$ and for any function $f: \mathbb{Z}^d \rightarrow \mathbb{R}$, we have*

$$\begin{aligned} (a) \quad & |\mathbb{E}\{\Delta f(W)^T b I_n^\delta(W)\} - n^{-1} \mathbb{E}\{(f(W)(W - nc)^T \Sigma^{-1} b) I_n^\delta(W)\}| \\ & \leq d^{1/2} C_{5.2}^{(1)}(\delta) n^{-1} \|b\|_1 \|f\|_{n\delta/2, \infty}^\Sigma; \\ (b) \quad & |\mathbb{E}\{\Delta f(W)^T B(W - nc) I_n^\delta(W)\} \\ & - \mathbb{E}\{f(W) [n^{-1}(W - nc)^T \Sigma^{-1} B(W - nc) - \text{Tr } B] I_n^\delta(W)\}| \\ & \leq d^{1/2} C_{5.2}^{(2)}(\delta) n^{-1/2} \|B\|_1 \|f\|_{n\delta/2, \infty}^\Sigma + \sum_{j=1}^d |B_{jj}| \|\Delta f\|_{n\delta/2, \infty}^\Sigma; \end{aligned}$$

$$\begin{aligned}
(c) \quad & |\mathbb{E}\{\Delta f(W)^T B(W - nc) I_n^\delta(W)\} \\
& - \mathbb{E}\{f(W) [n^{-1}(W - nc)^T \Sigma^{-1} B(W - nc) - \text{Tr } B] I_n^\delta(W)\}| \\
& \leq dC_{5.2}^{(3)}(\delta) n^{-1/2} \sum_{j=1}^d |(e^{(j)})^T B| \|f\|_{n\delta/2, \infty}^\Sigma + \sum_{j=1}^d |B_{jj}| \|\Delta f\|_{n\delta/2, \infty}^\Sigma,
\end{aligned}$$

for any d -vector b and any $d \times d$ matrix B . The constants $n_{5.2}$, $C_{5.2}^{(1)}(\delta)$ and $C_{5.2}^{(2)}(\delta)$ are defined in (7.50), (7.51) and (7.57), respectively.

With the help of these two lemmas, we can now show that, if W has the discrete normal distribution $\mathcal{DN}_d(nc, n\Sigma)$, then it satisfies the conditions of Theorem 4.7, with $\varepsilon_1 \leq c_1 n^{-1/2}$, $\max\{\varepsilon_{20}, \varepsilon_{21}\} \leq c_2 d^{5/2} n^{-1/2}$ and $\varepsilon_{22} = 0$.

Theorem 5.3. *For Σ positive definite, suppose that σ^2 , positive definite, and A are such that $A\Sigma + \Sigma A^T + \sigma^2 = 0$; write $\bar{\Lambda} := \bar{\lambda}(\sigma^2)$. Then, if $W \sim \mathcal{DN}_d(nc, n\Sigma)$, for any $n \geq \max\{n_{5.2}, \psi_\Sigma(\delta)\}$, we have*

$$\begin{aligned}
(i) \quad & \mathbb{E}|W - nc|_\Sigma^2 \leq dVn; \\
(ii) \quad & d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq C_{5.3}^{(1)} n^{-1/2}, \text{ for each } 1 \leq j \leq d; \\
(iii) \quad & |\mathbb{E}\{\mathcal{A}_n h(W) I_n^\delta(W)\}| \\
& \leq d^{5/2} n^{-1/2} \bar{\Lambda} C_{5.3}^{(2)}(\delta) (\|h\|_{n\delta/2, \infty}^\Sigma + n^{1/2} \|\Delta h\|_{n\delta/2, \infty}^\Sigma),
\end{aligned}$$

where \mathcal{A}_n is as defined in (1.1), $V = C(2)$ as in Lemma 5.1, and $C_{5.3}^{(1)}$ and $C_{5.3}^{(2)}(\delta)$ are functions of $\|A\|/\bar{\Lambda}$ and the elements of $\text{Sp}'(\Sigma)$ and $\text{Sp}'(\sigma^2/\bar{\Lambda})$; $C_{5.3}^{(1)}$ is given in (5.2), and $C_{5.3}^{(2)}$ implicitly in (5.8).

Proof. Part (i) is immediate from (7.43), with $V = C(2)$. For Part (ii), we pick $\delta = 1$, and then take $b = e^{(j)}$ and any function f with $\|f\|_\infty \leq 1$ in Lemma 5.2(a). This gives

$$\mathbb{E}|\Delta_j f(W) I_n^1(W)| \leq d^{1/2} C_{5.2}^{(1)}(1) n^{-1} + n^{-1/2} \sqrt{C'(1)(1 + (\Sigma^{-1})_{jj})},$$

in view of Lemma 5.1(c). For the remaining part of $|\mathbb{E}\{\Delta_j f(W)\}|$, using $\|\Delta_j f\|_\infty \leq 2$, we have

$$\mathbb{E}|\Delta_j f(W) I[|W - nc|_\Sigma > n/3]| \leq 18dC(2)/n,$$

by Chebyshev's inequality and from Part (i), and the estimate follows because $n \geq d^2$, with

$$C_{5.3}^{(1)} := C_{5.2}^{(1)}(1) + \sqrt{C'(1)(1 + (\Sigma^{-1})_{jj})} + 18C(2). \quad (5.2)$$

For Part (iii), we use Lemma 5.2(b). This gives

$$\begin{aligned}
& \left| \mathbb{E} \{ \Delta h(W)^T A(W - nc) I_n^\delta(W) \} \right. \\
& \quad \left. - \mathbb{E} \{ h(W) [n^{-1}(W - nc)^T \Sigma^{-1} A(W - nc) - \text{Tr } A] I_n^\delta(W) \} \right| \\
& \leq d^{1/2} C_{5.2}^{(2)}(\delta) n^{-1/2} \|A\|_1 \|h\|_{n\delta/2, \infty}^\Sigma + \sum_{j=1}^d |A_{jj}| \|\Delta h\|_{n\delta/2, \infty}^\Sigma. \quad (5.3)
\end{aligned}$$

Then, since

$$\text{Tr}(\sigma^2 \Delta^2 h(W)) = \sum_{i=1}^d \sum_{j=1}^d \sigma_{ij}^2 \Delta_j f_i(W),$$

where $f_i(W) := \Delta_i h(W)$, it follows from Lemma 5.2(a), with $f = f_i$ and with b the i -th column of σ^2 , that

$$\begin{aligned}
& \left| n \mathbb{E} \{ \text{Tr}(\sigma^2 \Delta^2 h(W)) I_n^\delta(W) \} \right. \\
& \quad \left. - \mathbb{E} \left\{ \sum_{i=1}^d \sum_{j=1}^d \sigma_{ij}^2 \Delta_i h(W) \{ \Sigma^{-1}(W - nc) \}_j I_n^\delta(W) \right\} \right| \\
& \leq d^{1/2} C_{5.2}^{(1)}(\delta) \|\sigma^2\|_1 \|\Delta h\|_{n\delta/2, \infty}^\Sigma; \quad (5.4)
\end{aligned}$$

note also that

$$\begin{aligned}
& \mathbb{E} \left\{ \sum_{i=1}^d \sum_{j=1}^d \sigma_{ij}^2 \Delta_i h(W) \{ \Sigma^{-1}(W - nc) \}_j I_n^\delta(W) \right\} \\
& = \mathbb{E} \{ \Delta h(W)^T \sigma^2 \Sigma^{-1}(W - nc) I_n^\delta(W) \}. \quad (5.5)
\end{aligned}$$

But now, from Lemma 5.2(c),

$$\begin{aligned}
& \left| \mathbb{E} \{ \Delta h(W)^T \sigma^2 \Sigma^{-1}(W - nc) I_n^\delta(W) \} \right. \\
& \quad \left. - \mathbb{E} \{ h(W) [n^{-1}(W - nc)^T \Sigma^{-1} \sigma^2 \Sigma^{-1}(W - nc) - \text{Tr}(\sigma^2 \Sigma^{-1})] I_n^\delta(W) \} \right| \\
& \leq d C_{5.2}^{(3)}(\delta) n^{-1/2} \sum_{j=1}^d |(e^{(j)})^T \Sigma^{-1} \sigma^2| \|h\|_{n\delta/2, \infty}^\Sigma + \sum_{j=1}^d |[\sigma^2 \Sigma^{-1}]_{jj}| \|\Delta h\|_{n\delta/2, \infty}^\Sigma \\
& \leq d C_{5.2}^{(3)}(\delta) n^{-1/2} \{\lambda_{\min}(\Sigma)\}^{-1} \|\sigma^2\|_1 \|h\|_{n\delta/2, \infty}^\Sigma + \|\sigma^2 \Sigma^{-1}\|_1 \|\Delta h\|_{n\delta/2, \infty}^\Sigma. \quad (5.6)
\end{aligned}$$

Hence, and since

$$\|A\|_1 \leq d^{3/2} \|A\|; \quad \|\sigma^2\|_1 \leq d^{3/2} \lambda_{\max}(\sigma^2)$$

and

$$\|\sigma^2 \Sigma^{-1}\|_1 \leq d^{3/2} \|\sigma^2 \Sigma^{-1}\| \leq d^{3/2} \lambda_{\max}(\sigma^2) / \lambda_{\min}(\Sigma),$$

it follows from (5.3), (5.4) and (5.6) that

$$\begin{aligned} & \mathbb{E}\{\mathcal{A}_n h(W) I_n^\delta(W)\} \\ &= \mathbb{E}\left\{(\text{Tr}\{A(W - nc)\Delta h(W)^T\} + \tfrac{1}{2}n \text{Tr}\{\sigma^2 \Delta^2 h(W)\}) I_n^\delta(W)\right\} \\ &= \mathbb{E}\left\{h(W) \left[\tfrac{1}{2}n^{-1}(W - nc)^T(2\Sigma^{-1}A + \Sigma^{-1}\sigma^2\Sigma^{-1})(W - nc) \right. \right. \\ &\quad \left. \left. - \text{Tr} A - \tfrac{1}{2}\text{Tr}(\sigma_\Sigma^2)\right] I_n^\delta(W)\right\} + \theta, \end{aligned} \quad (5.7)$$

where

$$\begin{aligned} |\theta| &\leq d^{1/2} C_{5.2}^{(2)}(\delta) n^{-1/2} \|A\|_1 \|h\|_{n\delta/2, \infty}^\Sigma \\ &\quad + \|A\|_1 \|\Delta h\|_{n\delta/2, \infty}^\Sigma + \tfrac{1}{2} d^{1/2} C_{5.2}^{(1)}(\delta) \|\sigma^2\|_1 \|\Delta h\|_{n\delta/2, \infty}^\Sigma \\ &\quad + \tfrac{1}{2} d C_{5.2}^{(3)}(\delta) n^{-1/2} \{\lambda_{\min}(\Sigma)\}^{-1} \|\sigma^2\|_1 \|h\|_{n\delta/2, \infty}^\Sigma + \|\sigma^2 \Sigma^{-1}\|_1 \|\Delta h\|_{n\delta/2, \infty}^\Sigma \\ &\leq d^{5/2} n^{-1/2} \bar{\Lambda} C_{5.3}^{(2)}(\delta) (\|h\|_{n\delta/2, \infty}^\Sigma + n^{1/2} \|\Delta h\|_{n\delta/2, \infty}^\Sigma), \end{aligned} \quad (5.8)$$

and $C_{5.3}^{(2)}(\delta)$ is a function of $\|A\|/\bar{\Lambda}$ and the elements of $\text{Sp}'(\Sigma)$, $\text{Sp}'(\sigma^2/\bar{\Lambda})$.

Finally, for any y and B , we have $y^T B y = y^T B^T y = \tfrac{1}{2} y^T (B + B^T) y$, so that

$$\begin{aligned} y^T (2\Sigma^{-1}A + \Sigma^{-1}\sigma^2\Sigma^{-1})y &= y^T (\Sigma^{-1}A + A^T\Sigma^{-1} + \Sigma^{-1}\sigma^2\Sigma^{-1})y \\ &= y^T \Sigma^{-1} (A\Sigma + \Sigma A^T + \sigma^2) \Sigma^{-1} y = 0, \end{aligned}$$

from (1.3), and

$$\text{Tr}(\sigma_\Sigma^2) = -\text{Tr}(\Sigma^{-1/2}A\Sigma^{1/2} + \Sigma^{1/2}A^T\Sigma^{-1/2}) = -2\text{Tr} A.$$

This, with (5.7), establishes that

$$\begin{aligned} & |\mathbb{E}\{\mathcal{A}_n h(W) I_n^\delta(W)\}| \\ &\leq d^{5/2} n^{-1/2} C_{5.3}^{(2)}(\delta) \left\{ \|h\|_{n\delta/2, \infty}^\Sigma + n^{1/2} \|\Delta h\|_{n\delta/2, \infty}^\Sigma \right\}, \end{aligned} \quad (5.9)$$

as required. \square

5.2 Discrete normal approximation by Stein's method

The aim now is to translate the conclusion of Theorem 5.3, which established properties of discrete normal distributions analogous to the conditions of Theorem 4.7, into a discrete normal approximation theorem. The first step is to relate a typical generator \mathcal{A}_n to the generator \mathcal{A}_n^δ of an appropriately chosen Markov population process. The existence of such a process, having the same local drift and covariance matrices, is a consequence of the following lemma, proved using Tropp (2015, Theorem 1.1).

Lemma 5.4. *Let σ^2 be any $d \times d$ covariance matrix with positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$. Then σ^2 can be represented in the form*

$$\sigma^2 = \sum_{J \in \mathcal{J}} \tilde{g}(J) J J^T,$$

for a finite set $\mathcal{J} \subset \mathbb{Z}^d$ such that $e^{(i)} \in \mathcal{J}$, $1 \leq i \leq d$, such that $J \in \mathcal{J}$ implies that $-J \in \mathcal{J}$, with $\tilde{g}(-J) = \tilde{g}(J)$, and such that

$$\max_{J \in \mathcal{J}} \max_{1 \leq i \leq d} |J_i| \leq 1 + \frac{1}{2} \sqrt{2(d-1)\rho(\sigma^2)}.$$

Furthermore, $\tilde{g}(e^{(i)}) \geq \frac{1}{4}\lambda_d$ for each $1 \leq i \leq d$.

Proof. Write $\lambda_0 := \frac{1}{2}\lambda_d = \frac{1}{2}\lambda_{\min}(\sigma^2)$, so that $\sigma^2 - \lambda_0 I$ is positive definite, and has condition number $\rho(\sigma^2 - \lambda_0 I) \leq 2\rho(\sigma^2)$. By Theorem 1.1 of Tropp (2015), we can write

$$\sigma^2 - \lambda_0 I = \sum_{J \in \mathcal{J}_1} \gamma(J) J J^T,$$

where the set \mathcal{J}_1 is finite, $\gamma(J) > 0$ for each $J \in \mathcal{J}_1$, and the vectors J have integer coordinates with $|J_i| \leq 1 + \frac{1}{2}\sqrt{(d-1)\rho(\sigma^2 - \lambda_0 I)}$. Note that the same covariance matrix is obtained if $\gamma(J) J J^T$ is replaced by $\frac{1}{2}\gamma(J)\{J J^T + (-J)(-J)^T\}$, which we do, expanding the set \mathcal{J}_1 if necessary. Writing $\lambda_0 I = \sum_{i=1}^d \frac{1}{2}\lambda_0\{e^{(i)}(e^{(i)})^T + (-e^{(i)})(-e^{(i)})^T\}$, and taking $\mathcal{J} = \mathcal{J}_1 \cup \{\pm e^{(i)}, 1 \leq i \leq d\}$, the lemma follows. \square

Using this lemma, we can now prove the normal approximation theorem analogous to Theorem 4.7. Note that $\psi_\Sigma(n) \leq \psi(n)$ for $n \geq n_\psi \in \mathcal{K}^{(3,1)}$, where

$$n_\psi := \left\{ \frac{3}{2} \sqrt{\frac{d\theta_1}{\lambda_{\min}(\Sigma)}} \right\}^{8/5}. \quad (5.10)$$

Theorem 5.5. Suppose that c is in \mathbb{R}^d , σ^2 is a $d \times d$ positive definite symmetric matrix and A a $d \times d$ matrix, all of whose eigenvalues have negative real parts. Let Σ be the positive definite solution to the equation $A\Sigma + \Sigma A^T + \sigma^2 = 0$, and let \mathcal{A}_n be as defined in (1.1). Write

$$\begin{aligned}\bar{\Lambda} &:= d^{-1} \text{Tr}(\sigma^2); \quad \delta_0 := \frac{\lambda_{\min}(\sigma^2)}{8\|A\|}; \\ \tilde{\delta}_0 &:= \frac{1}{\sqrt{\lambda_{\max}(\Sigma)}} \min\left\{ \frac{4\bar{\Lambda}\rho(\Sigma)}{\lambda_{\min}(\sigma^2)}, \frac{\delta_0}{3} \right\},\end{aligned}$$

and let

$$n_{5.5} := \max\{n_{5.2}, n_{3.6}, n_\psi, \psi^{-1}(\tilde{\delta}_0/2)\} \in \mathcal{K}^{(3.1)}.$$

Then, for any $V > 0$ and $\delta' \leq \tilde{\delta}_0/2$, there is a constant $C_{5.5}(V, \delta')$, depending only on $\text{Sp}'(\sigma^2/\bar{\Lambda})$, $\text{Sp}'(\Sigma)$, $\|A\|/\bar{\Lambda}$, V and δ' , with the following property: if W is a random vector in \mathbb{Z}^d such that, for some $\varepsilon_1, \varepsilon_{20}, \varepsilon_{21}, \varepsilon_{22}$, and for some $n \geq \max\{n_{5.5}, \psi^{-1}(\delta')\}$,

- (a) $\mathbb{E}|W - nc|_\Sigma^2 \leq dVn$;
- (b) $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon_1$, for each $1 \leq j \leq d$;
- (c) $|\mathbb{E}\{\mathcal{A}_n h(W)\} I[|W - nc|_\Sigma \leq n\delta'/3]|$
 $\leq \bar{\Lambda}(\varepsilon_{20}\|h\|_{n\tilde{\delta}_0/4, \infty}^\Sigma + \varepsilon_{21}n^{1/2}\|\Delta h\|_{n\tilde{\delta}_0/4, \infty}^\Sigma + \varepsilon_{22}n\|\Delta^2 h\|_{n\tilde{\delta}_0/4, \infty}^\Sigma),$

then

$$\begin{aligned}d_{\text{TV}}(\mathcal{L}(W), \mathcal{DN}_d(nc, n\Sigma)) \\ \leq C_{5.5}(V, \delta')(d^3 n^{-1/2} + d^{7/2}(\bar{\gamma}(\sigma^2)/\bar{\Lambda})\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22}) \log n;\end{aligned}$$

$\bar{\gamma}(\sigma^2)$ is as defined in (5.11) below, and $d^{-1/2}\bar{\gamma}(\sigma^2)/\bar{\Lambda}$ is bounded by a function of $\rho(\sigma^2/\bar{\Lambda})$ alone.

Remark 5.6. Note that $4\bar{\Lambda}\rho(\Sigma)/\lambda_{\min}(\sigma^2) = K_{3.5}/L_1$, relating the definition of $\tilde{\delta}_0$ to the definition of δ_1 in Theorem 3.6. Note also that $n_{5.5}$ is a function of $\text{Sp}'(\sigma^2/\bar{\Lambda})$, $\text{Sp}'(\Sigma)$ and $\|A\|/\bar{\Lambda}$ alone, in view of (5.12) and (5.13).

Proof. Represent σ^2 as in Lemma 5.4. For $J \in \mathcal{J}$, define

$$g^J(x) := \begin{cases} \tilde{g}(J), & \text{if } J \in \mathcal{J} \setminus \{e^{(i)}, 1 \leq i \leq d\}; \\ \tilde{g}(e^{(i)}) + (A(x - c))_i, & \text{for } 1 \leq i \leq d. \end{cases}$$

With these functions g^J , we have $\sigma^2 = \sum_{J \in \mathcal{J}} g^J(c) J J^T$ and, writing $F(x) := \sum_{J \in \mathcal{J}} J g^J(x)$, we also have $F(c) = 0$ and $DF(c) = A$; we further define

$$\bar{\gamma}(\sigma^2) = d^{-3/2} \sum_{J \in \mathcal{J}} g^J(c) |J|^3, \quad (5.11)$$

noting that, from Lemma 5.4, $d^{-1/2} \bar{\gamma}(\sigma^2) / \bar{\Lambda}$ is bounded above by a function of $\rho(\sigma^2 / \bar{\Lambda})$.

For δ_0 chosen as above, the quantity ε_0 in Assumption S4 is at least $1/2$. This is the case because all the transition rates $g^J(x)$ are constant in x , except for $J = e^{(i)}$, $1 \leq i \leq d$, when they are linear. For $g^{e^{(i)}}$, we have

$$\frac{g^{e^{(i)}}(x)}{g^{e^{(i)}}(c)} = \frac{\frac{1}{2} \tilde{g}(e^{(i)}) + (A(x - c))_i}{\frac{1}{2} \tilde{g}(e^{(i)})},$$

and this is at least $1/2$ if

$$|x - c| \|A\| \leq \frac{1}{8} \lambda_{\min}(\sigma^2) \leq \frac{1}{4} \tilde{g}(e^{(i)}),$$

which is in turn true if $|x - c| \leq \delta_0$. The same calculation shows that $L_0 \leq 3/2$, and it is also immediate, from Lemma 5.4, that

$$\begin{aligned} L_1 &\leq 2\|A\| / \min_{1 \leq i \leq d} \tilde{g}(e^{(i)}) \leq 4\|A\| / \lambda_{\min}(\sigma^2); \\ \bar{\Lambda} / g_* &\leq 2\bar{\Lambda} / \min_{1 \leq i \leq d} \tilde{g}(e^{(i)}) \leq 4\bar{\Lambda} / \lambda_{\min}(\sigma^2). \end{aligned} \quad (5.12)$$

Finally, again from Lemma 5.4,

$$d^{-1} J_{\max} \leq d^{-1} \{d(1 + \frac{1}{2} \sqrt{2(d-1)\rho(\sigma^2)})^2\}^{1/2} \leq 2 + \rho(\sigma^2 / \bar{\Lambda}). \quad (5.13)$$

Hence, for this choice of δ_0 , the quantities in $\mathcal{K}^{(3.1)}$ are all continuous functions of $\|A\| / \bar{\Lambda}$ and the elements of $\text{Sp}'(\sigma^2 / \bar{\Lambda})$ and $\text{Sp}'(\Sigma)$.

We now use the inequality

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{DN}_d(nc, n\Sigma)) \leq d_{\text{TV}}(\mathcal{L}(W), \Pi_n^{\tilde{\delta}_0}) + d_{\text{TV}}(\Pi_n^{\tilde{\delta}_0}, \mathcal{DN}_d(nc, n\Sigma))$$

where $\Pi_n^{\tilde{\delta}_0}$ denotes the equilibrium distribution of $X_n^{\tilde{\delta}_0}$, which is a Markov population process with transition rates given by $ng_{\tilde{\delta}_0}^J(n^{-1}X)$, where $g_{\tilde{\delta}_0}^J(x)$ is derived from $g^J(x)$ as in (1.10). From the definition of $\tilde{\delta}_0$ in the statement of the theorem, the conditions of Theorem 4.7 are satisfied, with L_0, L_1 and J_{\max} bounded as above. Since $\delta' \leq \tilde{\delta}_0/2$ and $n \geq \max\{n_{3.6}, \psi^{-1}(\delta')\}$, it follows from Theorem 4.7 that

$$d_{\text{TV}}(\mathcal{L}(W), \Pi_n^{\tilde{\delta}_0}) \leq C'_1 (d^3 n^{-1/2} + d^{7/2} (\bar{\gamma}(\sigma^2) / \bar{\Lambda}) \varepsilon_1 + \varepsilon_{20} + d^{1/4} \varepsilon_{21} + d^{1/2} \varepsilon_{22}) \log n,$$

for a constant C'_1 that depends on $V, \delta', \|A\|/\bar{\Lambda}$ and the elements of $\text{Sp}'(\sigma^2/\bar{\Lambda})$ and $\text{Sp}'(\Sigma)$, but not on n . On the other hand, from Theorem 5.3, if $\mathcal{L}(W) = \mathcal{DN}_d(nc, n\Sigma)$ and $n \geq \max\{n_{5.2}, n_{3.6}, \psi^{-1}(\delta')\}$, then Conditions (i)–(iii) of Theorem 4.7 are satisfied with $\delta' \leq \tilde{\delta}_0/2$, with $\varepsilon_1 \leq C_{5.3}^{(1)} n^{-1/2}$, with $\varepsilon_{22} = 0$, and with $\max\{\varepsilon_{20}, \varepsilon_{21}\} \leq C_{5.3}^{(2)}(\delta') d^{5/2} n^{-1/2}$. Hence, from Theorem 4.7,

$$d_{\text{TV}}(\mathcal{DN}_d(nc, n\Sigma), \Pi_n^{\tilde{\delta}_0}) \leq C'_2 d^{7/2} (\bar{\gamma}(\sigma^2)/\bar{\Lambda}) n^{-1/2} \log n,$$

for a constant C'_2 that depends on $V, \delta', \|A\|/\bar{\Lambda}$ and the elements of $\text{Sp}'(\sigma^2/\bar{\Lambda})$ and $\text{Sp}'(\Sigma)$. The triangle inequality completes the proof. \square

To check the conditions of the theorem, the presence of the norm $|\cdot|_{\Sigma}$ throughout is awkward, depending, as it does, on the covariance matrix Σ . It is more natural to have a theorem expressed instead in terms of the usual Euclidean norm in \mathbb{R}^d , and we frame our main Theorem 1.1 in such terms. We show in Appendix 7.5 that Theorem 1.1 is indeed implied by Theorem 5.5.

6 Examples

In this section, we give some applications of the total variation approximation theorem. The first is to the (quasi-)equilibrium distributions of a wide range of processes used as models in ecology and epidemiology. We then give two further classes of examples. The first is to sums of independent, integer valued random vectors, since this is the most obvious test case for any normal approximation. The second concerns random vectors for which there is an exchangeable pair, in the sense of Stein (1986, p.13). For both of these, we use a more general theorem, applicable in a setting in which there is a ‘linear regression pair.’

6.1 Markov population processes

Suppose that $(X_n, n \geq 1)$ is a fixed sequence of Markov population processes with $X_n(\cdot) \in n^{-1}\mathbb{Z}^d$ for some fixed d , and with transition rates determined by the fixed collection of functions $(g^J: \mathbb{R}^d \rightarrow \mathbb{R}_+, J \in \mathcal{J})$, satisfying Assumptions G0–G4 for some c and δ_0 . Then, for large n , X_n has a quasi-equilibrium behaviour near nc , in the sense that the process, if started near nc , remains within any ball $\tilde{B}_{n,\delta}(c)$ for a length of time whose expectation, for fixed $\delta > 0$, grows exponentially with n . During this time, its behaviour is asymptotically extremely close to that of X_n^δ , the choice of $\delta < \delta_0$ having almost no effect: see Barbour & Pollett (2012), Section 4. It thus has a quasi-equilibrium

distribution that, as $n \rightarrow \infty$, is asymptotically extremely close to Π_n^δ , for any $0 < \delta < \delta_0$. Here, we prove that Π_n^δ is itself close to a discrete normal distribution. This would follow directly from Theorems 4.7 and 5.3 if X_n satisfied Assumptions S2–S4. Here, since we only assume G2–G4, further argument is needed, in order to verify the assumptions of Theorem 5.5.

We begin by noting that the variance of Π_n^δ is of the correct order, satisfying Condition (a) of Theorem 5.5, and that Condition (b) is also satisfied, with $\varepsilon_1 = O(n^{-1/2})$. The proofs of these two results are in Appendix 7.7. Since, for this application, all the data of the problem, apart from n , are fixed, we can simplify many statements to order expressions as $n \rightarrow \infty$.

Lemma 6.1. *Let X_n be a Markov population process whose transition rates are given in (1.4), satisfying Assumptions G0–G4 for some $\delta_0 > 0$, and let $\delta_{2.2}$ and $\delta'_{2.2}(d)$ be as in Lemma 2.2. Then, for any $0 < \delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, if $X_n^\delta \sim \Pi_n^\delta$, we have*

$$\mathbb{E}|X_n^\delta - nc|_\Sigma^2 = O(n).$$

Proposition 6.2. *Under Assumptions G0–G4, if $X_n^\delta \sim \Pi_n^\delta$ for some fixed $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, then, for each $1 \leq j \leq d$,*

$$d_{\text{TV}}(\Pi_n^\delta, \Pi_n^\delta * \varepsilon_{e(j)}) = O(n^{-1/2}),$$

where ε_J denotes the point mass at J and $*$ denotes convolution.

Theorem 6.3. *Under the above assumptions, for any fixed $0 < \delta < \tilde{\delta}_0$,*

$$d_{\text{TV}}(\Pi_n^\delta, \mathcal{DN}_d(nc, n\Sigma)) = O(n^{-1/2} \log n),$$

as $n \rightarrow \infty$, where Σ satisfies $A\Sigma + \Sigma A^T + \sigma^2 = 0$, the matrices A and σ^2 are as given in (1.6) and (1.7), and $\tilde{\delta}_0$ is as for Theorem 5.5.

Proof. We show that, for $W \sim \Pi_n^\delta$, Conditions (a)–(c) of Theorem 5.5 are satisfied, with suitable choices of V , ε_1 and ε_{2l} , $0 \leq l \leq 2$. We use Theorem 5.5 rather than Theorem 1.1, because the main part of the proof is furnished by Theorem 4.6, which is expressed in the $|\cdot|_\Sigma$ norm.

Condition (a) follows immediately from Lemma 6.1, and Condition (b) is implied by Proposition 6.2, with $\varepsilon_1 = O(n^{-1/2})$. It then follows from Theorem 4.6 with $\delta' = \delta/2$ that, for any function h , we have

$$\begin{aligned} & |\mathbb{E}\{(\mathcal{A}_n^\delta h(W) - \mathcal{A}_n h(W))I[|W - nc|_\Sigma \leq n\delta/6]\}| \\ &= O(n^{-1/2}(n^{1/2}\|\Delta h\|_{n\delta/4, \infty}^\Sigma + n\|\Delta^2 h\|_{n\delta/4, \infty}^\Sigma)). \end{aligned} \quad (6.1)$$

However, since Π_n^δ is the equilibrium distribution of X_n^δ , it follows that $\mathbb{E}\{\mathcal{A}_n^\delta h(W)\} = 0$. Then, since $|W - nc|_\Sigma \leq n\delta$ implies that $|W - nc| \leq \delta_0$, because $\delta < \tilde{\delta}_0 \leq \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$, it follows that

$$\begin{aligned} & |\mathbb{E}\{\mathcal{A}_n^\delta h(W)I[|W - nc|_\Sigma > n\delta/6]\}| \\ & \leq n \sum_{J \in \mathcal{J}} |g^J|_{\delta_0} \mathbb{E}\{|h(W + J) - h(W)|I[|W - nc|_\Sigma > n\delta/6]\} \\ & \leq 2nL_0\Lambda \|h\| \mathbb{P}[|W - nc|_\Sigma > n\delta/6], \end{aligned} \tag{6.2}$$

where δ_0 , L_0 , Λ and $|\cdot|_\delta$ are as in Section 2.1. From Lemma 7.1, with $r = 2$, $\mathbb{P}[|W - nc|_\Sigma > n\delta/6] = O(n^{-2})$ as $n \rightarrow \infty$. Then, since the left hand side of (6.1) is unchanged if we set $h(X) = 0$ for $|X - nc|_\Sigma > n\delta/4$, provided that $J_{\max}^\Sigma \leq n\delta/12$, we can replace $\|h\|$ by $\|h\|_{n\delta/4, \infty}^\Sigma$ in (6.2) for all n sufficiently large. These two observations imply, with (6.2), that

$$|\mathbb{E}\{\mathcal{A}_n^\delta h(W)I[|W - nc|_\Sigma \leq n\delta/6]\}| = O(n^{-1}\|h\|_{n\delta/4, \infty}^\Sigma).$$

Combining this with (6.1), it follows that

$$\begin{aligned} & |\mathbb{E}\{\mathcal{A}_n h(W)I[|W - nc|_\Sigma \leq n\delta/6]\}| \\ & = O(n^{-1/2}(\|h\|_{n\delta/4, \infty}^\Sigma + n^{1/2}\|\Delta h\|_{n\delta/4, \infty}^\Sigma + n\|\Delta^2 h\|_{n\delta/4, \infty}^\Sigma)), \end{aligned}$$

which in turn implies that Condition (c) of Theorem 5.5 is satisfied, with $\varepsilon_{20}, \varepsilon_{21}$ and ε_{22} all of order $O(n^{-1/2})$, proving the result. \square

6.2 Linear regression pairs

In this section, we establish a discrete normal approximation theorem for the distribution of a random vector W , when a copy W' can be defined on the same probability space, in such a way that $\mathbb{E}\{W' | W\}$ is approximately a linear function of W . There are many examples where this is the case, including those given in Rinott & Rotar (1996) and Reinert & Röllin (2009). The theorem is easier to apply if the pair (W, W') is exchangeable, as discussed in Section 6.4 below.

Suppose, then, that (W, W') is a pair of random integer valued d -vectors, defined on the same probability space and having the same distribution. Assume that $\mathbb{E}\{|W|^3\} < \infty$, and write $\mu := \mathbb{E}W$. Let ξ denote the difference $W' - W$, so that $\mathbb{E}\xi = 0$, and set $\sigma^2 := \mathbb{E}\{\xi\xi^T\}$, assumed positive definite, and $\chi := \mathbb{E}\{|\xi|^3\}$. Suppose that ξ exhibits an almost linear regression on W , and that the conditional variance $\sigma^2(W) := \mathbb{E}\{\xi\xi^T | W\}$ is more or less constant as a function of W . Specifically, assume that, for some $n > 0$ and for some

$d \times d$ matrix A with spectral norm $\|A\|$, all of whose eigenvalues have negative real parts, we have

$$\begin{aligned}\mathbb{E}\{\xi \mid W\} &= n^{-1}A(W - \mu) + n^{-1/2}\|A\|^{1/2}R_1(W); \\ \sigma^2(W) &:= \mathbb{E}\{\xi\xi^T \mid W\} = \sigma^2 + R_2(W),\end{aligned}\tag{6.3}$$

where $\mathbb{E}|R_1(W)|$ and $\mathbb{E}\|R_2(W)\|_1$ are to be thought of as small. These two quantities appear explicitly in the bound on the error in our discrete normal approximation, and, clearly, $\mathbb{E}\{R_1(W)\} = 0$ and $\mathbb{E}\{R_2(W)\} = 0$. Let Σ be the positive definite solution to $A\Sigma + \Sigma A^T + \sigma^2 = 0$, and write $\alpha_1 := \frac{1}{2}\lambda_{\min}(\Sigma)$.

Remark 6.4. *Note that, in (6.3), multiplying n and A by the same positive constant c does not change the regression, but Σ is divided by c . This leaves both $n\Sigma$, the asymptotic approximation to $\text{Var } W$, and $\|A\|/n$ unchanged, the latter implying that $R_1(W)$ remains the same also. The effective data for the problem are the distributions of ξ and W , and in particular σ^2 and $\text{Var } W$, and also $\hat{A} := A/n$, which is typically ‘small’. In order to circumvent the indeterminacy, one can compute $\hat{\Sigma} := n\Sigma$, typically ‘large’, by solving $\hat{A}\hat{\Sigma} + \hat{\Sigma}\hat{A}^T + \sigma^2 = 0$. Then $\tilde{n} := n/\|A\|$, $\tilde{A} := \tilde{n}\hat{A} = A/\|A\|$ and $\tilde{\Sigma} := \hat{\Sigma}/\tilde{n}$ are the same for all c , yield the same regression matrix $\tilde{A}/\tilde{n} = \hat{A}$, and can be used as a standard version, if required.*

We now define further parameters

$$\begin{aligned}v &:= \text{Tr}(\sigma_\Sigma^2)/(d\alpha_1); \quad L := (\|A\|/n)^{1/2}\chi\{\text{Tr}(\sigma_\Sigma^2)\}^{-3/2}; \\ \chi_\Sigma &:= \mathbb{E}|\Sigma^{-1/2}\xi|^3; \quad L_\Sigma := (\|A\|/n)^{1/2}\chi_\Sigma\{\text{Tr}(\sigma_\Sigma^2)\}^{-3/2} \leq L\rho(\Sigma)^{3/2},\end{aligned}\tag{6.4}$$

and set $Z := z(W)$, where $z(w) := (ndv)^{-1/2}\Sigma^{-1/2}(w - \mu)$. L , L_Σ and Z all involve A , n and Σ only through the standardized quantities $n/\|A\|$ and $n\Sigma$. We then assume that the following inequalities hold:

$$\{\|A\|/\alpha_1\}^{1/2}\mathbb{E}\{(1 + |Z|)|\Sigma^{-1/2}R_1(W)|\} \leq \frac{1}{2}(\text{Tr}(\sigma_\Sigma^2))^{1/2}(1 + \mathbb{E}|Z|^2); \tag{6.5}$$

$$\{\|A\|/\alpha_1\}^{1/2}\mathbb{E}\{|Z|(1 + |Z|)|\Sigma^{-1/2}R_1(W)|\} \leq \frac{1}{4}(\text{Tr}(\sigma_\Sigma^2))^{1/2}(1 + \mathbb{E}|Z|^3). \tag{6.6}$$

They can reasonably be expected to be satisfied if $|R_1(W)|$ is indeed small. In particular, (6.5)–(6.6) are satisfied if

$$\{\mathbb{E}|\Sigma^{-1/2}R_1(W)|^3\}^{1/3} \leq \frac{1}{8}(\alpha_1\text{Tr}(\sigma_\Sigma^2)/\|A\|)^{1/2}. \tag{6.7}$$

Under the above conditions, the second and third moments of $|Z|$ can be suitably bounded; the proof is given in Appendix 7.6.

Lemma 6.5. *If (6.5) and (6.6) hold, and if $n/\alpha_1 \geq 1$, then*

$$\mathbb{E}|Z|^2 \leq 2; \quad \mathbb{E}|Z|^3 \leq m_3 := 2 \left(1 + \frac{10\chi_\Sigma}{(\text{Tr}(\sigma_\Sigma^2))^{3/2}} \right),$$

where $Z = z(W)$, with $z(w)$ as defined above. In particular, for any $\delta > 0$,

$$\frac{n}{\|A\|} \mathbb{P}[|W - \mu|_\Sigma > n\delta\|A\|^{-1/2}] \leq 2d^{3/2}\delta^{-3}((\|A\|/n)^{1/2} + 10L_\Sigma) \left\{ \frac{2\bar{\lambda}(\sigma_\Sigma^2)}{\lambda_{\min}(\sigma_\Sigma^2)} \right\}^{3/2}.$$

Remark 6.6. *Note that*

$$\left\{ |W - \mu|_\Sigma > \frac{n\delta}{\sqrt{\|A\|}} \right\} = \left\{ |Z| > \delta \sqrt{\frac{n}{\|A\|}} \sqrt{\frac{\lambda_{\min}(\sigma_\Sigma^2)}{2\bar{\lambda}(\sigma_\Sigma^2)}} \right\} \quad (6.8)$$

involves only standardized quantities.

We are now in a position to prove a discrete normal approximation theorem. To state it, we introduce some further notation:

$$\begin{aligned} \varepsilon_1 &:= \max_{1 \leq j \leq d} d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})); \\ \varepsilon_1(\xi) &:= \max_{1 \leq j \leq d} d_{\text{TV}}(\mathcal{L}(W | \xi), \mathcal{L}(W + e^{(j)} | \xi)). \end{aligned} \quad (6.9)$$

Theorem 6.7. *Assume that (W, W') is a pair of random integer valued d -vectors, such that $\mathcal{L}(W) = \mathcal{L}(W')$ and that $\mathbb{E}|W|^3 < \infty$; write $\mu := \mathbb{E}W$. Suppose that $\xi := W' - W$ satisfies the regression condition (6.3), for matrices A and σ^2 such that the eigenvalues of A all have negative real parts and σ^2 is positive definite; let Σ be the positive definite solution of $A\Sigma + \Sigma A^T + \sigma^2 = 0$. Define $\mathbb{E}|\xi|^3 := \chi$, $\alpha_1 := \frac{1}{2}\lambda_{\min}(\sigma_\Sigma^2)$, $\bar{\Lambda} := d^{-1}\text{Tr}(\sigma^2)$ and $L := (\|A\|/n)^{1/2}\chi\{\text{Tr}(\sigma^2)\}^{-3/2}$, and assume that (6.5) and (6.6) hold. Let $\bar{\gamma}(\sigma^2)$ be as for Theorem 5.5, and let \tilde{A} and $\tilde{\Sigma}$ be as in Remark 6.4. Then there exist constants n_0 and C , depending on $\|\tilde{A}\|$ and the elements of $\text{Sp}'(\sigma^2)$ and $\text{Sp}'(\tilde{\Sigma})$, such that, if $n/\|A\| \geq n_0$, we have*

$$\begin{aligned} & d_{\text{TV}}(\mathcal{L}(W), \mathcal{DN}_d(\mu, n\Sigma)) \\ & \leq C \log n \{ d^3(\|A\|/n)^{1/2} + d^{7/2}(\bar{\gamma}(\sigma^2)/\bar{\Lambda})\varepsilon_1 + d^{1/4}\mathbb{E}|R_1(W)| \\ & \quad + d^{1/2}\mathbb{E}\|R_2(W)\|_1 + d^3L + d^2\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\} \}. \end{aligned}$$

Proof. Because $\mathcal{L}(W) = \mathcal{L}(W')$, we have

$$\begin{aligned} 0 &= (n/\|A\|)\mathbb{E}\{h(W')I[|W' - \mu|_\Sigma \leq M] - h(W)I[|W - \mu|_\Sigma \leq M]\} \\ &= (n/\|A\|)\mathbb{E}\{(h(W') - h(W))I[|W - \mu|_\Sigma \leq M]\} \\ &\quad + (n/\|A\|)\mathbb{E}\{h(W')(I[|W' - \mu|_\Sigma \leq M] - I[|W - \mu|_\Sigma \leq M])\}, \end{aligned} \quad (6.10)$$

for any $M > 0$. We shall take $M = n\eta/6\sqrt{\|A\|}$, for η to be prescribed later, in view of (6.8). For bounded functions $h: \mathbb{Z}^d \rightarrow \mathbb{R}$, the second term can be simply estimated, using Lemma 6.5, by

$$\begin{aligned} \theta_0 &:= 2(n/\|A\|)\|h\|_\infty \mathbb{P}[|W - \mu|_\Sigma > M] \\ &\leq 864 d^{3/2} \eta^{-3} ((\|A\|/n)^{1/2} + 10L_\Sigma) \left\{ \frac{2\bar{\lambda}(\sigma_\Sigma^2)}{\lambda_{\min}(\sigma_\Sigma^2)} \right\}^{3/2} \|h\|_\infty. \end{aligned} \quad (6.11)$$

For the first term, we write

$$h(W') - h(W) = \xi^T \Delta h(W) + \frac{1}{2} \xi^T \Delta^2 h(W) \xi + e_2(W, \xi, h). \quad (6.12)$$

From (6.3), its first element yields

$$\begin{aligned} &\frac{n}{\|A\|} |\mathbb{E}\{\xi^T \Delta h(W) I[|W - \mu|_\Sigma \leq M]\} \\ &\quad - \mathbb{E}\{n^{-1}(W - \mu)^T A^T \Delta h(W) I[|W - \mu|_\Sigma \leq M]\}| \\ &\leq (n/\|A\|)^{1/2} \mathbb{E}\{|R_1(W)^T \Delta h(W)| I[|W - \mu|_\Sigma \leq M]\} \\ &\leq (n/\|A\|)^{1/2} \mathbb{E}|R_1(W)| \|\Delta h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty} =: \theta_1. \end{aligned} \quad (6.13)$$

Then

$$\begin{aligned} &\frac{n}{2\|A\|} |\mathbb{E}\{\xi^T \Delta^2 h(W) \xi I[|W - \mu|_\Sigma \leq M]\} \\ &\quad - \mathbb{E}\{\text{Tr}(\sigma^2 \Delta^2 h(W)) I[|W - \mu|_\Sigma \leq M]\}| \\ &\leq \frac{1}{2} \mathbb{E}\{\|R_2(W)\|_1\} (n/\|A\|) \|\Delta^2 h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty} =: \theta_2. \end{aligned} \quad (6.14)$$

It remains to bound $(n/\|A\|)\mathbb{E}\{e_2(W, \xi, h) I[|W - \mu|_\Sigma \leq M]\}$. We first consider $|\xi| > \sqrt{n/\|A\|}$, and use the bound

$$\mathbb{E}\{|\xi|_1^r I[|\xi| > \sqrt{n/\|A\|}]\} \leq d^{r/2} \mathbb{E}\{|\xi|^r I[|\xi| > \sqrt{n/\|A\|}]\} \leq d^{r/2} \chi(n/\|A\|)^{-(3-r)/2} \quad (6.15)$$

for $r = 0, 1, 2$. Since

$$\begin{aligned} &|e_2(W, \xi, h)| I[|W - \mu|_\Sigma \leq M] \\ &\leq 2\|h\|_\infty + |\xi|_1 \|\Delta h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty} + \frac{1}{2} |\xi|_1^2 \|\Delta^2 h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty}, \end{aligned}$$

it follows, using (6.15), that

$$\begin{aligned} \theta_3 &:= \frac{n}{\|A\|} \mathbb{E}\{|e_2(W, \xi, h)| I[|W - \mu|_\Sigma \leq M] I[|\xi| > \sqrt{n/\|A\|}]\} \\ &\leq \chi \sqrt{\frac{\|A\|}{n}} \left\{ 2\|h\|_\infty + \left(\frac{dn}{\|A\|} \right)^{1/2} \|\Delta h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty} + \frac{dn}{2\|A\|} \|\Delta^2 h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty} \right\} \\ &\leq 2L \{\text{Tr}(\sigma^2)\}^{3/2} \left\{ \|h\|_\infty + \left(\frac{dn}{\|A\|} \right)^{1/2} \|\Delta h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty} + \frac{dn}{\|A\|} \|\Delta^2 h\|_{\frac{\Sigma}{6\sqrt{\|A\|}}, \infty} \right\}. \end{aligned} \quad (6.16)$$

For $|\xi| \leq \sqrt{n/\|A\|}$, as in (4.10), we write

$$e_2(W, \xi, h) = E_2(W, \xi, h) - \frac{1}{2} \sum_{j=1}^d \xi_j \Delta_{jj} h(W). \quad (6.17)$$

For the contribution from the second term in (6.17), we have at most

$$\begin{aligned} \theta_4 &:= \frac{n}{2\|A\|} \left| \mathbb{E} \left\{ \sum_{j=1}^d \xi_j \Delta_{jj} h(W) I[|W - \mu|_\Sigma \leq M] I[|\xi| \leq \sqrt{n/\|A\|}] \right\} \right| \\ &\leq \frac{n}{2\|A\|} \left| \mathbb{E} \left\{ \sum_{j=1}^d \xi_j \Delta_{jj} h(W) I[|W - \mu|_\Sigma \leq M] \right\} \right| \\ &\quad + \frac{1}{2} \mathbb{E}\{|\xi|_1 I[|\xi| > \sqrt{n/\|A\|}]\} \frac{n}{\|A\|} \|\Delta^2 h\|_{\frac{n\eta}{6\sqrt{\|A\|}}, \infty}^\Sigma \\ &=: \theta'_4 + \theta''_4, \end{aligned} \quad (6.18)$$

say. Now, from (6.3) and (6.5), we have

$$\begin{aligned} \theta'_4 &= \frac{1}{2\|A\|} \sum_{j=1}^d \left| \mathbb{E} \left\{ ([A(W - \mu)]_j + n^{1/2} \|A\|^{1/2} [R_1(W)]_j) \Delta_{jj} h(W) I[|W - \mu|_\Sigma \leq M] \right\} \right| \\ &\leq \frac{1}{2} n^{-1/2} \left\{ (dv)^{1/2} \mathbb{E}|A \Sigma^{1/2} Z(W)|_1 + \|A\|^{1/2} \mathbb{E}|R_1(W)|_1 \right\} \frac{n}{\|A\|} \|\Delta^2 h\|_{\frac{n\eta}{6\sqrt{\|A\|}}, \infty}^\Sigma \\ &\leq \frac{1}{2} (\|A\|/n)^{1/2} d \sqrt{v} (\|A\|^{-1/2} \|\Sigma^{1/2}\|) \{\sqrt{2}\|A\| + \frac{3}{2}\alpha_1\} (n/\|A\|) \|\Delta^2 h\|_{\frac{n\eta}{6\sqrt{\|A\|}}, \infty}^\Sigma. \end{aligned} \quad (6.19)$$

Then, from (6.15),

$$\theta''_4 \leq \frac{1}{2} \{d^{1/2} (\|A\|/n) \chi\} (n/\|A\|) \|\Delta^2 h\|_{\frac{n\eta}{6\sqrt{\|A\|}}, \infty}^\Sigma. \quad (6.20)$$

For the first term in (6.17), by Lemma 4.4(i), if $|\xi| \leq \sqrt{n/\|A\|}$ and $n\eta/24\sqrt{\|A\|} \geq \sqrt{n/\|A\|} \lambda_{\min}(\Sigma)$, then

$$\begin{aligned} \theta_5(\xi) &:= (n/\|A\|) |\mathbb{E}\{E_2(W, \xi, h) I[|W - \mu|_\Sigma \leq M] \mid \xi\}| \\ &\leq \{C_{4,4}^{(1)}(\xi) \varepsilon_1(\xi) + C_{4,4}^{(2)}(\xi) \mathbb{P}[|W - \mu|_\Sigma \geq M/4 \mid \xi]\} (n/\|A\|) \|\Delta^2 h\|_{\frac{n\eta}{4\sqrt{\|A\|}}, \infty}^\Sigma \\ &\leq \{d^{3/2} |\xi|^3 \varepsilon_1(\xi) + 2d |\xi|^2 \mathbb{P}[|W - \mu|_\Sigma \geq M/4 \mid \xi]\} (n/\|A\|) \|\Delta^2 h\|_{\frac{n\eta}{4\sqrt{\|A\|}}, \infty}^\Sigma, \end{aligned}$$

where the last line uses Remark 4.5. Taking expectations, and then using

Lemma 6.5, this gives

$$\begin{aligned}
\theta_5 &:= \mathbb{E}\{|\theta_5(\xi)|I[|\xi| \leq \sqrt{n/\|A\|}]\} \\
&\leq \left\{ d^{3/2}\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\} + \frac{2dn}{\|A\|}\mathbb{P}[|W - \mu|_\Sigma \geq M/4] \right\} \frac{n}{\|A\|} \|\Delta^2 h\|_{\frac{n\eta}{4\sqrt{\|A\|}}, \infty}^\Sigma \\
&\leq \left\{ d^{3/2}\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\} + \eta^{-3}d^{5/2}C\{\rho(\sigma_\Sigma^2)\}^{3/2}\left(\sqrt{\frac{\|A\|}{n}} + L_\Sigma\right) \right\} \frac{n}{\|A\|} \|\Delta^2 h\|_{\frac{n\eta}{4\sqrt{\|A\|}}, \infty}^\Sigma,
\end{aligned} \tag{6.21}$$

for C a universal constant.

Let

$$\tilde{\mathcal{A}}_{\tilde{n}}h(w) := \frac{1}{2}\tilde{n}\text{Tr}(\sigma^2\Delta^2h(w)) + (w - \mu)^T \tilde{A}^T \Delta h(w). \tag{6.22}$$

Then, combining the estimates (6.11) and (6.13)–(6.21) with (6.10) and (6.12), we have shown that

$$\begin{aligned}
&|\mathbb{E}\{\tilde{\mathcal{A}}_{\tilde{n}}h(W)I[|W - \mu|_{\tilde{\Sigma}} \leq \tilde{n}\eta/6]\}| \\
&= |\frac{1}{2}\tilde{n}\mathbb{E}\{\text{Tr}(\sigma^2\Delta^2h(W))I[|W - \mu|_\Sigma \leq n\eta/6\sqrt{\|A\|}]\} \\
&\quad + \mathbb{E}\{(W - \mu)^T \tilde{A}^T \Delta h(W)I[|W - \mu|_\Sigma \leq n\eta/6\sqrt{\|A\|}]\}| \\
&\leq \sum_{l=0}^3 \theta_l + \theta'_4 + \theta''_4 + \theta_5 \\
&\leq \varepsilon_{20}\|h\|_\infty + \varepsilon_{21}\tilde{n}^{1/2}\|\Delta h\|_{\frac{n\eta}{4\sqrt{\|A\|}}, \infty}^\Sigma + \varepsilon_{22}\tilde{n}\|\Delta^2 h\|_{\frac{n\eta}{4\sqrt{\|A\|}}, \infty}^\Sigma \\
&\leq \bar{\Lambda}\{\varepsilon'_{20}\|h\|_\infty + \varepsilon'_{21}\tilde{n}^{1/2}\|\Delta h\|_{\frac{\tilde{\Sigma}}{n\eta/4}, \infty}^{\tilde{\Sigma}} + \varepsilon'_{22}\tilde{n}\|\Delta^2 h\|_{\frac{\tilde{\Sigma}}{n\eta/4}, \infty}^{\tilde{\Sigma}}\},
\end{aligned} \tag{6.24}$$

with

$$\begin{aligned}
\varepsilon'_{20} &= C_0(\eta)d^{3/2}(\tilde{n}^{-1/2} + L); & \varepsilon'_{21} &= \bar{\Lambda}^{-1}(\mathbb{E}|R_1(W)| + 2Ld^2\bar{\Lambda}^{3/2}); \\
\varepsilon'_{22} &= C_2(\eta)(\mathbb{E}\|R_2(W)\|_1 + Ld^{5/2} + d^{5/2}\tilde{n}^{-1/2} + d^{3/2}\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\}),
\end{aligned}$$

where the constants $C_l(\eta)$ depend on η , $\|\tilde{A}\|$ and the elements of $\text{Sp}'(\sigma^2)$ and $\text{Sp}'(\tilde{\Sigma})$. Since, if $\tilde{n}\eta/12 > 2/\sqrt{\lambda_{\min}(\tilde{\Sigma})}$, the quantity in (6.23) does not change if $h(X)$ is replaced by zero for $|X - \mu|_{\tilde{\Sigma}} > \tilde{n}\eta/4$, the norm $\|h\|_\infty$ can be replaced by $\|h\|_{\frac{\tilde{\Sigma}}{\tilde{n}\eta/4}, \infty}^{\tilde{\Sigma}}$ for such \tilde{n} and η . Thus Condition (c) of Theorem 5.5 is satisfied, for $\tilde{\mathcal{A}}_{\tilde{n}}$ as defined in (6.22), if we take $\eta = \tilde{\delta}_0$, for $\tilde{\delta}_0$ as defined in Theorem 5.5, and for \tilde{n} such that $\tilde{n} \geq \max\{n_{5.5}, 24/\{\tilde{\delta}_0\sqrt{\lambda_{\min}(\tilde{\Sigma})}\}\}$. The remaining conditions of Theorem 5.5, with $\tilde{\Sigma}$ for Σ and with \tilde{n} for n , are easily checked: Condition (a) is implied by Lemma 6.5, with $V = 2v$, and Condition (b) is just (6.9). This proves the theorem. \square

Remark 6.8. Direct computation of the quantities $\mathbb{E}|R_1(W)|$ and $\mathbb{E}\|R_2(W)\|_1$ can be awkward. However, it may be easier to find bounds on

$$\tilde{R}_1 := n^{1/2}\{\mathbb{E}(\xi | \mathcal{F}) - n^{-1}A(W - \mu)\} \quad \text{and} \quad \tilde{R}_2 := \mathbb{E}(\xi \xi^T | \mathcal{F}) - \sigma^2,$$

for a σ -field \mathcal{F} such that W is \mathcal{F} -measurable. From the properties of conditional expectation and Jensen's inequality, it follows that, for any non-negative random variable $Y(W)$, we have

$$\begin{aligned} \mathbb{E}\{Y(W)|R_1(W)|\} &\leq \mathbb{E}\{Y(W)|\tilde{R}_1|\}; \\ \mathbb{E}\{Y(W)\|R_2(W)\|_1\} &\leq \mathbb{E}\{Y(W)\|\tilde{R}_2\|_1\}. \end{aligned}$$

Hence we can use \tilde{R}_1 and \tilde{R}_2 in place of $R_1(W)$ and $R_2(W)$ when computing the bounds in the theorem and in verifying conditions (6.5)–(6.6).

In general, it may be difficult to find expressions bounding the total variation distance $d_{TV}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$, and its conditional versions. There are two circumstances in which this can be satisfactorily accomplished; when W is a sum of independent random variables, and when the pair (W, W') is *exchangeable*. These are discussed in the following sections.

6.3 Sums of independent integer valued random vectors

Let Y_i , $1 \leq i \leq m$, be independent \mathbb{Z}^d -valued random vectors, with means μ_i and covariance matrices S_i , and let $\gamma_i := \mathbb{E}|Y_i - \mu_i|^3$. Write $\mathbb{P}[Y_i = X] =: p_{i,X}$, $X \in \mathbb{Z}^d$, and define $u_i := \min_{1 \leq j \leq d} \{1 - d_{TV}(\mathcal{L}(Y_i), \mathcal{L}(Y_i + e^{(j)}))\}$. Let

$$\begin{aligned} W &:= \sum_{i=1}^m Y_i; \quad \mu := \mathbb{E}W = \sum_{i=1}^m \mu_i; \quad s_m := \sum_{i=1}^m u_i; \\ S &:= \mathbb{E}\{(W - \mu)(W - \mu)^T\} = \sum_{i=1}^m S_i; \quad \Gamma := \sum_{i=1}^m \gamma_i. \end{aligned}$$

We now apply Theorem 6.7 to approximate the distribution of W .

To start with, we need to define a W' on the same probability space, in such a way that $\mathcal{L}(W') = \mathcal{L}(W)$, and such that $\xi = W' - W$ is not too large. The canonical way to do this (Stein, 1986, p.16) is to let (Y'_1, \dots, Y'_m) be an independent copy of (Y_1, \dots, Y_m) , and to let K be uniformly distributed on $\{1, 2, \dots, m\}$, independently of the Y_i and the Y'_i ; then W' is taken to

be $W - Y_K + Y'_K$. It is clear that $\mathcal{L}(W') = \mathcal{L}(W)$, and also, writing $\xi := W' - W = Y'_K - Y_K$, that

$$\begin{aligned}\mathbb{E}(\xi | W) &= \mathbb{E}\{\mathbb{E}(\xi | Y_1, \dots, Y_m) | W\} \\ &= \mathbb{E}\left\{m^{-1} \sum_{i=1}^m (\mu_i - Y_i) \middle| W\right\} = -m^{-1}(W - \mu),\end{aligned}$$

so that the regression condition in (6.3) is satisfied with $A/n = -I/m$, and with $R_1(W) = 0$. Then $\sigma^2 = \mathbb{E}\{\xi\xi^T\} = 2S/m$, giving, for the standardized quantities of Remark 6.4, $\Sigma_0 = S$ and $\alpha_0 = 1/m$, and hence $\tilde{n} = m$, $\tilde{A} = -I$ and $\tilde{\Sigma} = S/m$. Note also that

$$\chi = \mathbb{E}|\xi|^3 = m^{-1} \sum_{i=1}^m \mathbb{E}|Y_i - Y'_i|^3 \leq 4m^{-1}\Gamma.$$

As a next step in applying Theorem 6.7, we show that the quantity ε_1 of (6.9) can be suitably bounded.

Lemma 6.9. *For W as defined above,*

$$\varepsilon_1 := \max_{1 \leq j \leq d} d_{TV}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) = O(s_m^{-1/2}).$$

Proof. Fix any $1 \leq j \leq d$, and, for $X \in \mathbb{Z}^d$, define

$$p_{i,X}^- := \frac{1}{2}(p_{i,X} \wedge p_{i,X-e^{(j)}}); \quad p_{i,X}^+ := \frac{1}{2}(p_{i,X} \wedge p_{i,X+e^{(j)}}).$$

Then define the pair (Y_i, \tilde{Y}_i) jointly, for $1 \leq i \leq m$, by

$$(Y_i, \tilde{Y}_i) = \begin{cases} (X, X - e^{(j)}) & \text{with probability } p_{i,X}^-; \\ (X, X + e^{(j)}) & \text{with probability } p_{i,X}^+; \\ (X, X) & \text{with probability } p_{i,X} - p_{i,X}^- - p_{i,X}^+, \end{cases} \quad X \in \mathbb{Z}^d.$$

Set $Z_i := Y_i - \tilde{Y}_i$. Then Z_i takes the values $e^{(j)}$ and $-e^{(j)}$ each with probability $\sum_{X \in \mathbb{Z}^d} p_{i,X}^+$, and takes the value 0 with probability $1 - \sum_{X \in \mathbb{Z}^d} p_{i,X} \wedge p_{i,X+e^{(j)}}$. Hence, for $T_0 := 0$ and $T_k := \sum_{i=1}^k Z_i$, the process $\{T_k, 0 \leq k \leq m\}$ is a lazy symmetric random walk. Define

$$Y'_i := \begin{cases} \tilde{Y}_i, & i \leq \tau; \\ Y_i, & i > \tau, \end{cases}$$

where $\tau := \min\{k: 1 \leq k \leq m, T_k = e^{(j)}\}$ if this is defined, and with $\tau = m$ otherwise. Set $W' = \sum_{i=1}^m Y'_i$. Then, by the Mineka coupling argument (Lindvall, 2002, Section II.14), it follows that

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \mathbb{P}[W \neq W' + e^{(j)}] \leq \mathbb{P}[\tau > m] = O(s_m^{-1/2}).$$

□

As a result of this lemma, it is clear that the quantity ε_1 of (6.9) is of order $O(s_m^{-1/2})$. Defining $W^{(i)} := W - Y_i$ and $\tilde{s}_m := s_m - \max_{1 \leq i \leq m} u_i$, we now observe that, for any $X \in \mathbb{Z}^d$, the conditional quantity $\varepsilon_1(X)$ is bounded by

$$\tilde{\varepsilon}_1 := \max_{1 \leq i \leq m} \max_{1 \leq j \leq d} d_{\text{TV}}(\mathcal{L}(W^{(i)}), \mathcal{L}(W^{(i)} + e^{(j)})) = O((\tilde{s}_m)^{-1/2}), \quad (6.25)$$

with the final order statement following directly from Lemma 6.9. This is because, for any $X \in \mathbb{Z}^d$,

$$\begin{aligned} & d_{\text{TV}}(\mathcal{L}(W + e^{(j)} | \xi = X), \mathcal{L}(W | \xi = X)) \\ & \leq m^{-1} \sum_{i=1}^m d_{\text{TV}}(\mathcal{L}(W + e^{(j)} | \xi = X, K = i), \mathcal{L}(W | \xi = X, K = i)), \end{aligned}$$

and because, by independence,

$$\begin{aligned} & d_{\text{TV}}(\mathcal{L}(W + e^{(j)} | \xi_i = X), \mathcal{L}(W | \xi_i = X)) \\ & \leq \mathbb{E}\{d_{\text{TV}}(\mathcal{L}(W^{(i)} + \xi_i + e^{(j)} | \xi_i, \xi'_i = \xi_i + X), \mathcal{L}(W^{(i)} + \xi_i | \xi_i, \xi'_i = \xi_i + X))\} \\ & = \mathbb{E}\{d_{\text{TV}}(\mathcal{L}(W^{(i)} + \xi_i + e^{(j)} | \xi_i), \mathcal{L}(W^{(i)} + \xi_i | \xi_i))\} \\ & = \mathbb{E}\{d_{\text{TV}}(\mathcal{L}(W^{(i)} + e^{(j)} | \xi_i), \mathcal{L}(W^{(i)} | \xi_i))\} \\ & = \mathbb{E}\{d_{\text{TV}}(\mathcal{L}(W^{(i)} + e^{(j)}), \mathcal{L}(W^{(i)}))\} \leq \tilde{\varepsilon}_1. \end{aligned}$$

Thus a number of the elements appearing in the bound given in Theorem 6.7 can be successfully handled. We now show that a multivariate discrete normal approximation can indeed be established. We write

$$\bar{\Lambda} := d^{-1} \text{Tr}(\sigma^2) = 2 \text{Tr}(S/m) \quad \text{and} \quad L := m^{-1/2} \frac{\chi}{\{\text{Tr}(\sigma^2)\}^{3/2}} \geq m^{-1/2};$$

the latter quantity, introduced in (6.4), is of order $O(m^{-1/2})$ if the ratio $\mathbb{E}|\xi|^3 / \{\mathbb{E}|\xi|^2\}^{3/2}$ remains bounded.

Theorem 6.10. *Under the above circumstances,*

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{DN}_d(\mu, S)) \leq C d^{7/2} \log m (L + m^{-1/2} \bar{\gamma}(S/m)) \sqrt{\frac{m}{\tilde{s}_m}},$$

for a suitable constant C , depending only on $\text{Sp}'(S/m)$.

Proof. With the definitions of W' and W given above, the regression condition in (6.3) is satisfied with $R_1(w) = 0$ for all $w \in \mathbb{Z}^d$, so that Conditions

(6.5) and (6.6) are trivially satisfied. Then $\varepsilon_1 = O(s_m^{-1/2})$, by Lemma 6.9, and

$$\mathbb{E}\{|\xi|^3 \varepsilon_1(\xi)\} = O((\tilde{s}_m)^{-1/2} \chi), \quad (6.26)$$

from the observations above. Note that

$$\chi = L\sqrt{m} d^{3/2} \overline{\Lambda}^{3/2}. \quad (6.27)$$

For $\mathbb{E}\|R_2(W)\|_1$, for any $X, w \in \mathbb{Z}^d$, we write $p(X) := \mathbb{P}[\xi = X]$, obtaining

$$\sigma_{il}^2(w) = \sum_{X \in \mathbb{Z}^d} X_i X_l \mathbb{P}[\xi = X \mid W = w] = \sum_{X \in \mathbb{Z}^d} p(X) X_i X_l \frac{\mathbb{P}[W = w \mid \xi = X]}{\mathbb{P}[W = w]}.$$

Hence

$$\begin{aligned} \mathbb{E}|\sigma_{il}^2(W) - \sigma_{il}^2| &= \sum_{w \in \mathbb{Z}^d} \left| \sum_{X \in \mathbb{Z}^d} p(X) X_i X_l (\mathbb{P}[W = w \mid \xi = X] - \mathbb{P}[W = w]) \right| \\ &= \sum_{w \in \mathbb{Z}^d} \left| \sum_{X \in \mathbb{Z}^d} p(X) X_i X_l \sum_{y \in \mathbb{Z}^d} p(y) (\mathbb{P}[W = w \mid \xi = X] - \mathbb{P}[W = w \mid \xi = y]) \right| \\ &\leq \sum_{X \in \mathbb{Z}^d} p(X) |X_i| |X_l| \sum_{y \in \mathbb{Z}^d} p(y) 2d_{\text{TV}}(\mathcal{L}(W \mid \xi = X), \mathcal{L}(W \mid \xi = y)). \end{aligned} \quad (6.28)$$

Now, by independence,

$$\begin{aligned} d_{\text{TV}}(\mathcal{L}(W \mid \xi = X), \mathcal{L}(W \mid \xi = y)) &\leq \frac{1}{m} \sum_{i=1}^m d_{\text{TV}}(\mathcal{L}(W^{(i)} + Y_i \mid Y'_i - Y_i = x), \mathcal{L}(W^{(i)} + Y_i \mid Y'_i - Y_i = y)) \\ &\leq \frac{1}{m} \sum_{i=1}^m \mathbb{E}\{d_{\text{TV}}(\mathcal{L}(W^{(i)} + Y_i \mid Y'_i, Y_i = Y'_i - x), \mathcal{L}(W^{(i)} + Y_i \mid Y'_i, Y_i = Y'_i - y))\} \\ &= \frac{1}{m} \sum_{i=1}^m \mathbb{E}\{d_{\text{TV}}(\mathcal{L}(W^{(i)}), \mathcal{L}(W^{(i)} - y + x))\} \\ &\leq \tilde{\varepsilon}_1 |y - x|_1. \end{aligned}$$

Substituting this bound into (6.28) and adding over $1 \leq i, l \leq d$ thus gives

$$\begin{aligned} \mathbb{E}\|R_2(W)\|_1 &\leq 2 \sum_{i=1}^d \sum_{l=1}^d \sum_{X \in \mathbb{Z}^d} p(X) |X_i| |X_l| \sum_{y \in \mathbb{Z}^d} p(y) \tilde{\varepsilon}_1 |x - y|_1 \\ &\leq 2\tilde{\varepsilon}_1 \sum_{X \in \mathbb{Z}^d} p(X) |X|_1^2 \{|X|_1 + \mathbb{E}|\xi|_1\} \\ &\leq 4\tilde{\varepsilon}_1 \mathbb{E}|\xi|_1^3 \leq 4\tilde{\varepsilon}_1 d^{3/2} \chi. \end{aligned} \quad (6.29)$$

It only remains to collect the elements needed for Theorem 6.7. It is immediate that $\bar{\gamma}(\sigma^2)/\bar{\Lambda} = \bar{\gamma}(S/m)$. Then, from (6.26) and (6.29), and from the definition of L , we have

$$d^{1/2}\mathbb{E}\|R_2(W)\|_1 + d^2\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\} = O(d^2\chi\tilde{s}_m^{-1/2}) = O(Ld^{7/2}(m/\tilde{s}_m)^{1/2}\bar{\Lambda}^{3/2}).$$

Combining this with the remaining elements of the bound given in Theorem 6.7, and noting that $\tilde{s}_m \leq m$, the theorem follows. \square

Except for the logarithmic factors, the bound obtained in the theorem is of the same order in m as would be expected for weaker metrics, such as the convex sets metric (Bentkus 2003, Fang & Röllin 2014), if $\tilde{s}_m \asymp m$. The latter asymptotic equivalence is what would be the case, for example, for identically distributed summands whose common distribution has non-trivial overlap with its unit translates in each direction. It is possible, however, for \tilde{s}_m to be significantly smaller than m . For instance, if all the summands making up W are on $2\mathbb{Z} \times \mathbb{Z}^{d-1}$, then $s_m = 0$, and the discrete normal is not a good approximation to W in total variation, since it puts about half its probability mass on points whose first coordinate is an odd integer, whereas $\mathcal{L}(W)$ puts zero mass on this set.

The best approximation order with respect to the convex sets metric, for sums of independent and identically distributed random variables with finite third moment, is $O(d^{7/4}L)$. Thus our rate is weaker in m by a factor of $\log m$, and in dimension by a factor of $d^{7/4}$, if $\bar{\gamma}(S/m)/\bar{\Lambda}$ is bounded. If the distributions are not identical, the best known d -dependence for approximation in the convex sets metric is rather worse, unless the random variables are also assumed to be bounded. Since the total variation metric is substantially stronger than the convex sets metric, our bounds are of encouragingly small order. Note also that all the key terms contributing to the error bound arise from estimates of the supremum of the quantity $\|\Delta^2 h(X)\|_\infty$ in appropriate balls, using Theorem 4.1. As noted in Remark 4.2, if A is a multiple of the identity, as in this example, the general bounds given in that theorem can be improved by removing the factor $d^{1/2}$. As a result, the dimension dependence in the statement of Theorem 6.10 can in fact be improved a little, replacing $d^{7/2}$ by d^3 .

6.4 Exchangeable pairs

If the pair (W, W') is also *exchangeable*, so that $\mathcal{L}((W, W')) = \mathcal{L}((W', W))$, a neat argument of Röllin & Ross (2015) delivers bounds on the quantities ε_1 and $\varepsilon_1(\xi)$ of (6.9), which appear in the bound given in Theorem 6.7. These

can be of considerable practical use in deriving explicit bounds from the general expressions given in Theorem 6.7.

Let \mathcal{J} be the set of d -vectors such that $q^J := \mathbb{P}[\xi = J] > 0$, and suppose that Assumption G4 of Section 2.1 is satisfied. For $Q^J(W) := \mathbb{P}[\xi = J | W]$, we set

$$u^J := (q^J)^{-1} \mathbb{E}|Q^J(W) - q^J|, \quad (6.30)$$

to be thought of as small. Note that, by exchangeability,

$$q^J = \mathbb{E}\{I[W' - W = J]\} = \mathbb{E}\{I[W - W' = J]\} = q^{-J}. \quad (6.31)$$

We then write

$$\tilde{u}_j := \sum_{l=1}^{r(j)} (u^{J_l^{(j)}} + u^{-J_l^{(j)}}),$$

where the $(J_l^{(j)}, 1 \leq l \leq r(j), 1 \leq j \leq d)$ are as in Assumption G4, and then set $\tilde{u}^* := \max_{1 \leq j \leq d} \tilde{u}_j$ and $u^* := \sup_{J \in \mathcal{J}} u^J$. With the help of these quantities, we can bound the differences $d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ between the distribution of W and its translates.

Lemma 6.11. *For each $1 \leq j \leq d$, we have*

$$d_{\text{TV}}(\mathcal{L}(W + e^{(j)}), \mathcal{L}(W)) \leq \tilde{u}_j,$$

and

$$d_{\text{TV}}(\mathcal{L}(W + e^{(j)} | \xi = J), \mathcal{L}(W | \xi = J)) \leq \tilde{u}_j + 2u^J.$$

Hence, in particular, for each $J \in \mathcal{J}$,

$$d_{\text{TV}}(\mathcal{L}(W + e^{(j)} | \xi = J), \mathcal{L}(W | \xi = J)) \leq \tilde{u}^* + 2u^*,$$

and $d_{\text{TV}}(\mathcal{L}(W + e^{(j)}), \mathcal{L}(W)) \leq \tilde{u}^*$. Furthermore, for $R_2(W)$ as defined in (6.3), we have

$$\mathbb{E}\|R_2(W)\|_1 \leq d \text{Tr}(\sigma^2) u^*.$$

Proof. For any $J \in \mathcal{J}$ and any f with $\|f\|_\infty = 1$, we use exchangeability to give

$$\mathbb{E}\{f(W')I[W' - W = J] - f(W)I[W - W' = J]\} = 0.$$

Dividing by q^J , using (6.31), and evaluating the expectation by conditioning on W , we obtain

$$\begin{aligned} 0 &= (q^J)^{-1} \mathbb{E}\{f(W + J)Q^J(W) - f(W)Q^{-J}(W)\} \\ &= \mathbb{E}\{f(W + J) - f(W)\} + (q^J)^{-1} \mathbb{E}\{f(W + J)(Q^J(W) - q^J)\} \\ &\quad - (q^{-J})^{-1} \mathbb{E}\{f(W)(Q^{-J}(W) - q^{-J})\}, \end{aligned}$$

from which it follows that

$$d_{\text{TV}}(\mathcal{L}(W + J), \mathcal{L}(W)) \leq u^J + u^{-J}.$$

The first statement now follows by the triangle inequality.

For the second, we have

$$\begin{aligned} & \mathbb{E}\{f(W + e^{(j)}) - f(W) \mid \xi = J\} \\ &= (q^J)^{-1} \mathbb{E}\{(f(W + e^{(j)}) - f(W))I[\xi = J]\} \\ &= (q^J)^{-1} \mathbb{E}\{(f(W + e^{(j)}) - f(W))Q^J(W)\} \\ &= \mathbb{E}\{f(W + e^{(j)}) - f(W)\} \\ &\quad + (q^J)^{-1} \mathbb{E}\{(f(W + e^{(j)}) - f(W))(Q^J(W) - q^J)\}. \end{aligned}$$

Hence we have

$$\begin{aligned} & d_{\text{TV}}(\mathcal{L}(W \mid \xi = J), \mathcal{L}(W + e^{(j)} \mid \xi = J)) \\ & \leq d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) + 2u^J, \end{aligned} \tag{6.32}$$

and the second part follows; note that exchangeability was not used in proving (6.32).

Finally, from the definition of $R_2(W)$ in (6.3), we have

$$\{R_2(w)\}_{il} = \sigma_{il}^2(w) - \sigma_{il}^2 = \sum_{J, J' \in \mathcal{J}} J_i J'_l (Q^J(w) - q^J),$$

for any $1 \leq i, l \leq d$, so that

$$\mathbb{E}|\sigma_{il}^2(W) - \sigma_{il}^2| \leq \sum_{J, J' \in \mathcal{J}} q^J |J_i| |J'_l| u^J \leq \mathbb{E}\{|\xi_i| |\xi_l|\} u^*.$$

This in turn implies that

$$\mathbb{E}\|R_2(W)\|_1 = \sum_{i=1}^d \sum_{l=1}^d \mathbb{E}|\sigma_{il}^2(W) - \sigma_{il}^2| \leq \mathbb{E}\{|\xi|_1^2\} u^* \leq d \text{Tr}(\sigma^2) u^*,$$

as claimed. \square

The following corollary is immediate.

Corollary 6.12. *Under the above assumptions,*

$$d^{1/2} \mathbb{E}\|R_2(W)\|_1 \leq C \bar{\Lambda} \tilde{n}^{-1/2} d^{5/2} \{\tilde{n}^{1/2} u^*\}$$

and

$$d^2 \mathbb{E}\{|\xi|^3 \varepsilon_1(\xi)\} \leq C' \bar{\Lambda}^{3/2} d^{7/2} L \{\tilde{n}^{1/2} (\tilde{u}^* + 2u^*)\},$$

for constants C and C' that depend only on $\text{Sp}'(\sigma^2/\bar{\Lambda})$. \square

Remark 6.13. Note that, by the argument in Remark 6.8, we can bound the quantities u^J above by $(q^J)^{-1}\mathbb{E}|\mathbb{P}[\xi = J | \mathcal{F}] - q^J|$, for any σ -field such that W is \mathcal{F} -measurable. Such quantities may be easier to bound in practice.

Remark 6.14. For an exchangeable pair (W, W') , we see that

$$\begin{aligned}\mathbb{E}\{\xi\xi^T\} &= \mathbb{E}\{(W' - \mu)(W' - W)^T - (W - \mu)(W' - W)^T\} \\ &= -\mathbb{E}\{-(W - \mu)(W - W')^T + (W - \mu)(W' - W)^T\} \\ &= -2\mathbb{E}\{(W - \mu)(W' - W)^T\} = -2\mathbb{E}\{(W - \mu)\mathbb{E}(\xi^T | W)\} \\ &= -2\mathbb{E}\{\mathbb{E}(\xi | W)(W - \mu)^T\},\end{aligned}\tag{6.33}$$

the last equality following because $\mathbb{E}\{\xi\xi^T\}$ is symmetric. If the remainders $R_1(W)$ and $R_2(W)$ in (6.3) were exactly zero, this would give

$$\frac{1}{2}\sigma^2 = -n^{-1}A \text{Cov}(W) = -n^{-1}\text{Cov}(W)A^T,$$

and hence also $A\sigma^2 = \sigma^2A^T$. If this is the case, we can easily solve for Σ , since then $\Sigma := -\frac{1}{2}A^{-1}\sigma^2 = -\frac{1}{2}\sigma^2(A^T)^{-1}$ satisfies $A\Sigma + \Sigma A^T + \sigma^2 = 0$ and is symmetric.

6.4.1 Monochrome edges in regular graphs

As an example of the application of Theorem 6.7 in the exchangeable setting, suppose that G_n is an r -regular graph on n vertices (so that one of n and r is even); thus there are $nr/2$ edges in the graph. Let the vertices be coloured independently, each with one of m colours, the probability of choosing colour i being $p_i > 0$, $1 \leq i \leq m$. Let N_i denote the number of vertices having colour i , and let M_i denote the number of edges joining pairs of vertices that both have colour i . We are interested in approximating the joint distribution of

$$W := (M_1, \dots, M_m, N_1, \dots, N_{m-1}) := (W_1, \dots, W_m, W_{m+1}, \dots, W_{2m-1}),$$

when n becomes large, while r , m and p_1, \dots, p_m remain fixed; the detailed structure of G_n does not appear in the approximation. Of course, the value of $N_m = n - \sum_{i=1}^{m-1} N_i$ is implied by knowledge of W . This problem, in the context of multivariate normal approximation, was considered by Rinott & Rotar (1996) and in Chen, Goldstein & Shao (2011, pp.333–334).

Theorem 6.15. For $m \geq 3$, r and p_1, \dots, p_m fixed, we can find $\nu \in \mathbb{R}^{2m-1}$ and a $(2m-1) \times (2m-1)$ covariance matrix Σ such that, as $n \rightarrow \infty$,

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{DN}_{2m-1}(n\nu, n\Sigma)) = O(n^{-1/2} \log n).$$

Proof. We use the notation of Theorem 6.7 throughout. We begin by observing that

$$\mathbb{E}M_i = nrp_i^2/2; \quad \mathbb{E}N_i = np_i,$$

determining $\nu := n^{-1}\mathbb{E}W$. After rather more calculation, the covariances are given, for $1 \leq i \neq l \leq m$, by

$$\begin{aligned} \text{Var}(M_i) &= \frac{1}{2}nrp_i^2(1-p_i)\{1+(2r-1)p_i\}; & \text{Cov}(N_i, N_l) &= -np_i p_l; \\ \text{Cov}(M_i, M_l) &= -\frac{1}{2}nr(2r-1)p_i^2 p_l^2; & \text{Cov}(M_i, N_l) &= -nrp_i^2 p_l; \\ \text{Cov}(M_i, N_i) &= nrp_i^2(1-p_i); & \text{Var}(N_i) &= np_i(1-p_i), \end{aligned} \tag{6.34}$$

in turn determining Σ .

We now construct an exchangeable pair (W, W') by first realizing a colouring $(C(j), 1 \leq j \leq n)$, and using it to define

$$M_i := \sum_{\{j, j'\} \in G} I[C(j) = C(j') = i] \quad \text{and} \quad N_i := \sum_{j=1}^n I[C(j) = i], \tag{6.35}$$

for each $1 \leq i \leq m$, thus defining W . We then choose a vertex K uniformly at random, independently of $(C(j), 1 \leq j \leq m)$, and then replace $C(K)$ by C' , where C' is independently sampled from $1, 2, \dots, m$ with $\mathbb{P}[C' = i] = p_i$, $1 \leq i \leq m$. If this new colouring is denoted by $(C'(j), 1 \leq j \leq m)$, then we define M'_i and N'_i as in (6.35), but with the $C'(j)$ in place of $C(j)$, and hence deduce W' . Of course, $\mathcal{L}(W, W') = \mathcal{L}(W', W)$, and W' differs from W only through the (possibly) new colour at the vertex K , and through its impact in changing which edges incident to K are monochrome:

$$\begin{aligned} M'_i - M_i &= \sum_{j: \{j, K\} \in G} (I[C(j) = C'(K) = i] - I[C(j) = C(K) = i]) \\ N'_i - N_i &= \{I[C'(K) = i] - I[C(K) = i]\}. \end{aligned}$$

Hence, for $1 \leq l \leq m$, we have

$$\begin{aligned} &\mathbb{E}\{\xi_l \mid C(1), \dots, C(n)\} \\ &= n^{-1} \sum_{k=1}^n \sum_{j: \{j, k\} \in G} \{p_l I[C(k) = l] - I[C(j) = C(k) = l]\} \\ &= n^{-1} \{p_l r N_l - 2M_l\} = \mathbb{E}\{\xi_l \mid W\}, \end{aligned}$$

and, for $m+1 \leq l \leq 2m-1$,

$$\mathbb{E}\{\xi_l \mid C(1), \dots, C(n)\} = n^{-1} \{np_{l-m} - N_{l-m}\} = \mathbb{E}\{\xi_l \mid W\}.$$

This gives an exact linear regression as in (6.3), with $R_1(w) = 0$ for all w , and with A having non-zero elements given by

$$\begin{aligned} A_{ll} &:= -2, & A_{l,l+m} &:= rp_l, & 1 \leq l \leq m-1; \\ A_{mm} &:= -2, & A_{m,m+t} &:= -rp_m, & 1 \leq t \leq m-1; \\ A_{ll} &:= -1, & m+1 \leq l \leq 2m-1. \end{aligned}$$

Since A is upper triangular, its eigenvalues are -2 , with multiplicity m , and -1 , with multiplicity $m-1$, so that it is indeed spectrally negative.

The set \mathcal{J} , consisting of the possible values that can be taken by ξ , is finite, and does not depend on n . If $C(K) = i \neq l = C'(K)$, then the $m+i$ and $m+l$ components of ξ each have modulus one (though, if i or l are equal to m , one of these components is not present in W), and the i and l components are in modulus at most r ; all other components of ξ are zero. Hence $|\xi|^2 \leq 2(r^2 + 1)$ a.s., and $\mathbb{E}|\xi|^3$ remains bounded as n increases; L is thus of strict order $n^{-1/2}$. The components of $\sigma^2 := \mathbb{E}\{\xi\xi^T\}$ can be explicitly calculated: for $1 \leq l \neq l' \leq m$, they are given by

$$\begin{aligned} \mathbb{E}\xi_l^2 &= 2p_l^2(1-p_l)\{r(r-1)p_l + r\}; & \mathbb{E}\{\xi_l\xi_{l'}\} &= -2r(r-1)p_l^2p_{l'}^2; \\ \mathbb{E}\{\xi_l\xi_{m+l}\} &= 2rp_l^2(1-p_l); & \mathbb{E}\{\xi_l\xi_{m+l'}\} &= -2rp_l^2p_{l'}; \\ \mathbb{E}\{\xi_{m+l}^2\} &= 2p_l; & \mathbb{E}\{\xi_{m+l}\xi_{m+l'}\} &= -2p_lp_{l'}, \end{aligned}$$

where terms with subscript $2m$ are to be ignored.

In order to apply Theorem 6.7, we now just need to find bounds for ε_1 , $\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\}$ and $\mathbb{E}\|R_2(W)\|_1$. From Lemma 6.11 and Corollary 6.12, these are all bounded by fixed multiples of u^* and \tilde{u}^* . For each J in the fixed finite set \mathcal{J} , the probability q^J in the denominator of u^J is fixed and positive, and hence bounded away from zero. To bound the numerator, we condition on a larger σ -field \mathcal{F} , with respect to which W is measurable, as in Remark 6.13. Let $T_{m,r}$ denote the set of all m -tuples of nonnegative integers t_1, \dots, t_m such that $\sum_{i=1}^m t_i = r$, and, for $t := (t_1, \dots, t_m) \in T_{m,r}$, let $E_j(i_0; t)$ denote the event that $C(j) = i_0$, and that t_i of the r neighbours of j have colour i , $1 \leq i \leq m$. For each fixed j , these are disjoint events whose union over $1 \leq i_0 \leq m$ and $t \in T_{m,r}$ is the sure event. We let \mathcal{F} be the σ -field generated by the events

$$\{E_j(i_0; t); 1 \leq j \leq n, 1 \leq i_0 \leq m, t \in T_{m,r}\}.$$

Then, if $K = j$, the value $J \in \mathcal{J}$ taken by ξ is determined by which of the events $(E_j(i_0; t); 1 \leq i_0 \leq m, t \in T_{m,r})$ occurs. For each J , there is a collection $S(J)$ of possible choices, consisting of just one possible $i_0 = i_0(J)$,

the index for which $J_{m+i_0} = -1$ (if there is none, then $i_0 = m$), but of all t that satisfy $t_{i_0} = -J_{i_0}$ and $t_{i_1} = J_{i_1}$, where i_1 is the index for which $J_{m+i_1} = 1$ (or m , if there is none such). Thus

$$\mathbb{P}[\xi = J | \mathcal{F}] = n^{-1} \sum_{j=1}^n \sum_{t \in T_{m,r} : (i_0(J), t) \in S(J)} I[E_j(i_0(J); t)].$$

Now, if $j' \neq j$ is such that the set of neighbours $\mathcal{N}(j)$ (including j) in G is disjoint from the set $\mathcal{N}(j')$, the events $I[E_j(i_0(J); t)]$ and $I[E_{j'}(i_0(J); t')]$ are independent. Since, for each j , there are no more than $r + r^2$ choices of $j' \neq j$ for which this is not the case, it follows that

$$\text{Var} \{ \mathbb{P}[\xi = J | \mathcal{F}] \} = O(n^{-1}).$$

Hence $\text{Var} \{ Q^J(W) \} = O(n^{-1})$ also, and so $\mathbb{E}|Q^J(W) - q^J| = O(n^{-1/2})$ for all $J \in \mathcal{J}$, implying that $u^* = O(n^{-1/2})$.

The argument for \tilde{u}^* is not yet finished, since, for each $1 \leq l \leq 2m - 1$, it is necessary to find a chain $J^{(1)}, J^{(2)}, \dots, J^{(R)}$ such that each $J^{(i)} \in \mathcal{J}$ and $\sum_{i=1}^R J^{(i)} = e^{(l)}$. For $m + 1 \leq l \leq 2m - 1$, this is easy: $\xi = e^{(l)}$ if, when W is constructed, a vertex has colour m and no neighbours of colours m or l , and its colour is replaced by l when resampling to obtain W' . Note that, to do this, we need at least three colours: $m \geq 3$. To get $e^{(l)}$ for $1 \leq l \leq m - 1$, a chain of length 2 is needed: a vertex of colour m with no neighbours of colour m and with exactly one of colour l is recoloured with colour l , giving $J = e^{(l)} + e^{(l+m)}$. Then $J = -e^{(l+m)}$ can be attained by reversing the order of the choices in the example for $m + 1 \leq l \leq 2m - 1$. To get $e^{(m)}$, a vertex of colour $l \neq m$ with no neighbours of colour l and exactly one of colour m is recoloured m , yielding $e^{(m)} - e^{(m+l)}$, and then adding $e^{(m+l)}$ as before completes the chain. Thus, for $m \geq 3$, we have $\tilde{u}^* = O(n^{-1/2})$ also, and applying Theorem 6.7, the result follows. \square

There remains the case of $m = 2$. Here, discrete normal approximation in total variation is not good, since it can be seen that $M_1 - M_2 = r(N_1 - n/2)$, so that W is degenerate; what is more, reducing to (W_1, W_2) gives an integer vector living on a proper sub-lattice of \mathbb{Z}^2 . However, the pair (M_1, N_1) can be approximated using the method above, and the remaining components of M and N follow from $N_2 = n - N_1$ and $M_2 = M_1 - r(N_1 - n/2)$.

7 Appendix

7.1 Proof of Lemma 4.3

In order to bound $|\mathbb{E}\{(f(W + X + U) - f(W + X))I[|W - nc|_\Sigma \leq n\delta/3]\}|$, we write

$$\begin{aligned} & (f(W + X + U) - f(W + X))I[|W - nc|_\Sigma \leq n\delta/3] \\ &= \tilde{f}_X(W + U) - \tilde{f}_X(W) \\ & \quad + f(W + X + U)\{I[|W - nc|_\Sigma \leq n\delta/3] - I[|W + U - nc|_\Sigma \leq n\delta/3]\}, \end{aligned}$$

where $\tilde{f}_X(Y) := f(Y + X)I[|Y - nc|_\Sigma \leq n\delta/3]$. Since $|\tilde{f}_X(Y)| \leq \|f\|_{n\delta/2, \infty}^\Sigma$ if $|X|_\Sigma \leq n\delta/6$, it is immediate that

$$|\mathbb{E}\{\tilde{f}_X(W + U) - \tilde{f}_X(W)\}| \leq \varepsilon_1 |U|_1 \|f\|_{n\delta/2, \infty}^\Sigma.$$

Then, on the set

$$\{|Y - nc|_\Sigma \leq n\delta/3 < |Y + U - nc|_\Sigma\} \cup \{|Y + U - nc|_\Sigma \leq n\delta/3 < |Y - nc|_\Sigma\},$$

it follows that $|Y - nc|_\Sigma > n\delta/3 - |U|_\Sigma \geq n\delta/4$ if $n\delta \geq 12|U|_\Sigma$, and that

$$|Y + X + U - nc|_\Sigma \leq n\delta/3 + \max\{|X|_\Sigma, |X + U|_\Sigma\} \leq n\delta/2,$$

this last by assumption. Hence

$$\begin{aligned} & |\mathbb{E}\{f(W + X + U)(I[|W - nc|_\Sigma \leq n\delta/3] - I[|W + U - nc|_\Sigma \leq n\delta/3])\}| \\ & \leq \varepsilon_2 \|f\|_{n\delta/2, \infty}^\Sigma, \end{aligned}$$

and Lemma 4.3 is proved. \square

7.2 Proof of Lemma 4.4

We prove only part (i), showing that, if $E_2(W, J, h)$ is as in (4.10), then

$$\begin{aligned} & \left| \mathbb{E}\left\{E_2(W, J, h)I[|W - nc|_\Sigma \leq n\delta/3]\right\} \right| \\ & \leq \|\Delta^2 h\|_{n\delta/2, \infty}^\Sigma (C_{4.4}^{(1)}(J)\varepsilon_1 + C_{4.4}^{(2)}(J)\varepsilon_2), \end{aligned}$$

for constants $C_{4.4}^{(1)}(J), C_{4.4}^{(2)}(J)$ given in (4.11), whenever

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon_1, \quad 1 \leq j \leq d; \quad \mathbb{P}[|W - nc|_\Sigma \geq n\delta/4] \leq \varepsilon_2;$$

the proof of part (ii) is entirely similar. We begin by taking any function $f: \mathbb{Z}^d \rightarrow \mathbb{R}$ and any $k \in \mathbb{Z}$ and $1 \leq j \leq d$. First, we note that, for $X \in \mathbb{Z}^d$,

$$f(X + ke^{(j)}) - f(X) = \begin{cases} \sum_{l=1}^k \Delta_j f(X + (l-1)e^{(j)}) & \text{if } k \geq 1; \\ -\sum_{l=1}^{|k|} \Delta_j f(X - le^{(j)}) & \text{if } k \leq -1. \end{cases} \quad (7.36)$$

Hence, by considering positive and negative k separately, we find that

$$|f(X + ke^{(j)}) - f(X) - k\Delta_j f(X)| \leq \frac{1}{2}|k|(|k| + 1)\|\Delta^2 f\|_\infty.$$

For more general increments $J \in \mathbb{Z}^d$, we define

$$J^{(s)} := (J_1, J_2, \dots, J_s, 0, 0, \dots, 0), \quad s \geq 1; \quad J^{(0)} := (0, \dots, 0).$$

Then, from the inequalities above, we have

$$|f(X + J^{(s)}) - f(X + J^{(s-1)}) - J_s \Delta_s f(X + J^{(s-1)})| \leq \frac{1}{2}|J_s|(|J_s| + 1)\|\Delta^2 f\|_\infty,$$

and

$$|\Delta_s f(X + J^{(s-1)}) - \Delta_s f(X)| \leq |J^{(s-1)}|_1 \|\Delta^2 f\|_\infty.$$

Hence it follows that

$$\begin{aligned} & |f(X + J^{(s)}) - f(X + J^{(s-1)}) - J_s \Delta_s f(X)| \\ & \leq \left\{ \frac{1}{2}|J_s|(|J_s| + 1) + |J_s||J^{(s-1)}|_1 \right\} \|\Delta^2 f\|_\infty. \end{aligned}$$

Adding over $1 \leq s \leq d$, this gives

$$|f(X + J) - f(X) - Df(X)^T J| \leq \frac{1}{2}|J|_1(|J|_1 + 1)\|\Delta^2 f\|_\infty.$$

The same argument also shows that

$$|f(X + J) - f(X) - Df(X)^T J| \leq \frac{1}{2}|J|_1(|J|_1 + 1)\|\Delta^2 f\|_{n\delta/2, \infty}^\Sigma, \quad (7.37)$$

if $|X - nc|_\Sigma \leq n\delta/3$ and $|J|_\Sigma \leq n\delta/6$.

We now prove Part (i) of the lemma by induction on the number r of non-zero components of J . We write $I_n^\eta(X)$ as shorthand for $I[|X - nc|_\Sigma \leq n\eta/3]$, for any $\eta > 0$. Starting with $r = 1$, we consider three cases. For $J = ke^{(j)}$ and $k \geq 1$, we have

$$\begin{aligned} & h(X + ke^{(j)}) - h(X) - k\Delta_j h(X) - \frac{1}{2}k(k-1)\Delta_{jj}h(X) \\ & = \sum_{l=1}^{k-1} \{\Delta_j h(X + le^{(j)}) - \Delta_j h(X)\} - \frac{1}{2}k(k-1)\Delta_{jj}h(X) \\ & = \sum_{l=1}^{k-1} \sum_{r=1}^{l-1} \{\Delta_{jj}h(X + re^{(j)}) - \Delta_{jj}h(X)\}. \end{aligned} \quad (7.38)$$

From Lemma 4.3, with $X = 0$ and $U = re^{(j)}$, it follows that

$$|\mathbb{E}\{(\Delta_{jj}h(W + re^{(j)}) - \Delta_{jj}h(W)) I_n^\delta(W)\}| \leq (r\varepsilon_1 + \varepsilon_2) \|\Delta^2 h\|_{n\delta/2, \infty}^\Sigma \quad (7.39)$$

for $r \leq k - 2$, if $|ke^{(j)}|_\Sigma = |J|_1|e^{(j)}|_\Sigma \leq n\delta/12$. Multiplying (7.38) by $I_n^\delta(X)$, replacing X by W , then taking expectations, invoking (7.39), and adding, this yields the claim for $J = ke^{(j)}$ and $k \geq 1$, with the upper bounds $C_{4.4}^{(1)}(ke^{(j)}) = \frac{1}{6}(k-2)(k-1)k$ and $C_{4.4}^{(2)}(ke^{(j)}) = \frac{1}{2}(k-1)k$. If $J = ke^{(j)}$ and $k = 0$, there is nothing to prove. For $J = -ke^{(j)}$ and $k \geq 1$, we have

$$\begin{aligned} & h(X - ke^{(j)}) - h(X) - (-k)\Delta_j h(X) - \frac{1}{2}(-k)(-k-1)\Delta_{jj}h(X) \\ &= \sum_{l=1}^k \{\Delta_j h(X) - \Delta_j h(X - le^{(j)})\} - \frac{1}{2}k(k+1)\Delta_{jj}h(X) \\ &= \sum_{l=1}^k \sum_{r=1}^l \{\Delta_{jj}h(X - re^{(j)}) - \Delta_{jj}h(X)\}. \end{aligned} \quad (7.40)$$

Arguing as before yields the claim for $J = -ke^{(j)}$ and $k \geq 1$, with

$$C_{4.4}^{(1)}(-ke^{(j)}) = \frac{1}{6}k(k+1)(k+2); \quad C_{4.4}^{(2)}(-ke^{(j)}) = \frac{1}{2}k(k+1),$$

again if $|ke^{(j)}|_\Sigma = |J|_1|e^{(j)}|_\Sigma \leq n\delta/12$. This establishes that the inequality (i) is true for $r = 1$, when J has just one non-zero component.

Now, for any $2 \leq r \leq d$, we assume that (i) is true for all J with at most $r-1$ non-zero components, and show that this implies that (i) is also true for all J with at most r non-zero components. Without loss of generality, we consider any J with $J_j = 0$ for $r < j \leq d$. First, we write

$$h(X + J) - h(X) = \{h(X + J) - h(X + J^{(r-1)})\} + \{h(X + J^{(r-1)}) - h(X)\}.$$

The induction hypothesis gives

$$\begin{aligned} & \left| \mathbb{E}\left\{ \left(e_2(W, J^{(r-1)}, h) + \frac{1}{2} \sum_{j=1}^{r-1} J_j \Delta_{jj} h(W) \right) I_n^\delta(W) \right\} \right| \\ & \leq \|\Delta^2 h\|_{n\delta/2, \infty}^\Sigma (C_{4.4}^{(1)}(J^{(r-1)})\varepsilon_1 + C_{4.4}^{(2)}(J^{(r-1)})\varepsilon_2). \end{aligned}$$

Thus it remains only to consider the expectation of the quantity

$$\begin{aligned} & (h(W + J) - h(W + J^{(r-1)}) - J_r \Delta_r h(W) \\ & - \frac{1}{2} J_r^2 \Delta_{rr} h(W) - J_r \sum_{j=1}^{r-1} J_j \Delta_{rj} h(W) + \frac{1}{2} J_r \Delta_{rr} h(W)) I_n^\delta(W). \end{aligned}$$

The one dimensional result gives

$$\begin{aligned} & \left| \mathbb{E} \left\{ \left(h(W + J) - h(W + J^{(r-1)}) - J_r \Delta_r h(W + J^{(r-1)}) \right. \right. \right. \\ & \quad \left. \left. \left. - \frac{1}{2} J_r (J_r - 1) \Delta_{rr} h(W + J^{(r-1)}) \right) I_n^\delta(W) \right\} \right| \\ & \leq \frac{1}{6} |J_r| (|J_r| + 1) \|\Delta^2 h\|_{n\delta/2, \infty}^{\Sigma} ((|J_r| + 2)\varepsilon_1 + 3\varepsilon_2), \end{aligned}$$

leaving an expectation involving the expression

$$\begin{aligned} J_r \left\{ \Delta_r h(X + J^{(r-1)}) - \Delta_r h(X) - \sum_{j=1}^{r-1} J_j \Delta_{rj} h(X) \right\} \\ + \frac{1}{2} J_r (J_r - 1) \{ \Delta_{rr} h(X + J^{(r-1)}) - \Delta_{rr} h(X) \}. \end{aligned} \quad (7.41)$$

The first line in (7.41) can be expressed as

$$J_r \sum_{s=1}^{r-1} \{ \Delta_r h(X + J^{(s)}) - \Delta_r h(X + J^{(s-1)}) - J_s \Delta_{rs} h(X) \}. \quad (7.42)$$

From (7.36) with $f := \Delta_d h$, we have

$$\begin{aligned} & \Delta_r h(X + J^{(s)}) - \Delta_r h(X + J^{(s-1)}) - J_s \Delta_{rs} h(X) \\ & = \begin{cases} \sum_{l=1}^{J_s} \{ \Delta_{rs} h(X + J^{(s-1)} + (l-1)e^{(s)}) - \Delta_{rs} h(X) \}, & \text{if } J_s \geq 1; \\ 0, & \text{if } J_s = 0; \\ - \sum_{l=1}^{|J_s|} \{ \Delta_{rs} h(X + J^{(s-1)} - le^{(s)}) - \Delta_{rs} h(X) \}, & \text{if } J_s \leq -1. \end{cases} \end{aligned}$$

Hence, multiplying by $I_n^\delta(X)$, taking expectations with W in place of X and using Lemma 4.3, it follows for the first line in (7.41) that

$$\begin{aligned} & |J_r| \left| \mathbb{E} \left\{ \sum_{s=1}^{r-1} \{ \Delta_r h(W + J^{(s)}) - \Delta_r h(W + J^{(s-1)}) - J_s \Delta_{rs} h(W) \} I_n^\delta(W) \right\} \right| \\ & \leq |J_r| \sum_{s=1}^{r-1} |J_s| \|\Delta^2 h\|_{n\delta/2, \infty}^{\Sigma} \{ (|J^{(s-1)}|_1 + \frac{1}{2}|J_s|)\varepsilon_1 + \varepsilon_2 \} \\ & \leq |J_r| |J^{(r-1)}|_1 \|\Delta^2 h\|_{n\delta/2, \infty}^{\Sigma} (\frac{1}{2}|J^{(r-1)}|_1 \varepsilon_1 + \varepsilon_2). \end{aligned}$$

The second line in (7.41) is directly bounded using Lemma 4.3, giving

$$\begin{aligned} & \frac{1}{2} |J_r| (|J_r| + 1) \left| \mathbb{E} \{ (\Delta_{rr} h(W + J^{(r-1)}) - \Delta_{rr} h(W)) I_n^\delta(W) \} \right| \\ & \leq \frac{1}{2} |J_r| (|J_r| + 1) \|\Delta^2 h\|_{n\delta/2, \infty}^{\Sigma} (|J^{(r-1)}|_1 \varepsilon_1 + \varepsilon_2). \end{aligned}$$

This establishes the inequality (i) for J with $J_j = 0$ for $r < j \leq d$, since it is easily checked that

$$C_{4.4}^{(1)}(J) \geq C_{4.4}^{(1)}(J^{(r-1)}) + \frac{1}{2}|J_r| |J^{(r-1)}|_1^2 + \frac{1}{2}|J_r|(|J_r| + 1) |J^{(r-1)}|_1,$$

and that

$$C_{4.4}^{(2)}(J) \geq C_{4.4}^{(2)}(J^{(r-1)}) + |J_r| |J^{(r-1)}|_1 + \frac{1}{2}|J_r|(|J_r| + 1).$$

The lemma now follows by induction. \square

7.3 Proof of Lemma 5.1

Let φ_n denote the density of the multivariate normal distribution $\mathcal{N}_d(nc, n\Sigma)$, and, for $X \in \mathbb{Z}^d$, let $[X]$ denote the box

$$[X] := \{x \in \mathbb{R}^d: X_i - \frac{1}{2} < x_i \leq X_i + \frac{1}{2}, 1 \leq i \leq d\}.$$

Let N_d denote a random vector having the standard d -dimensional normal distribution. For (a), the bound on $\mathbb{E}|W - nc|_\Sigma^l$ is obtained by first writing

$$|X - nc|_\Sigma^l \leq (|X - t|_\Sigma + |t - nc|_\Sigma)^l \leq 2^{l-1}(|X - t|_\Sigma^l + |t - nc|_\Sigma^l).$$

Taking this inside the integral, we have

$$\begin{aligned} \mathbb{E}|W - nc|_\Sigma^l &= \sum_{X \in \mathbb{Z}^d} |X - nc|_\Sigma^l \int_{[X]} \varphi_n(t) dt \\ &\leq \sum_{X \in \mathbb{Z}^d} \int_{[X]} \varphi_n(t) 2^{l-1} ((\frac{1}{2}\sqrt{d/\lambda_{\min}(\Sigma)})^l + |t - nc|_\Sigma^l) dt \\ &\leq \mathbb{E}\{2^{l-1}((\frac{1}{2}\sqrt{d/\lambda_{\min}(\Sigma)})^l + n^{l/2}|N_d|^l)\} \\ &\leq 2^l \mathbb{E}|N_d|^l n^{l/2}, \end{aligned}$$

where $N_d \sim \mathcal{N}_d(0, I)$, for

$$n \geq \frac{d}{4(\mathbb{E}|N_d|)^2 \lambda_{\min}(\Sigma)} = \frac{d}{8\lambda_{\min}(\Sigma)} \{\Gamma(d/2)/\Gamma((d+1)/2)\}^2.$$

Part (a) follows, taking $C(l) := 2^l \sqrt{\mathbb{E}|N_1|^{2l}}$, since

$$2^l \mathbb{E}|N_d|^l \leq 2^l \sqrt{\mathbb{E}|N_d|^{2l}} \leq 2^l d^{l/2} \sqrt{\mathbb{E}|N_1|^{2l}},$$

and by noting that $\frac{d}{8} \{\Gamma(d/2)/\Gamma((d+1)/2)\}^2 \leq 1$ in $d \geq 1$.

For (c), the bound on $\mathbb{E}\{[\Sigma^{-1}(W - nc)]_j^{2l}\}$, we first note, using very rough estimates, that

$$\mathbb{E}\{(a^T N_d)^{2l}\} \leq (a^T a)^l \mathbb{E}\{N_1^{2l}\} \frac{(2l)!}{2^l l!} = (a^T a)^l \left(\frac{(2l)!}{2^l l!} \right)^2 =: k(l)(a^T a)^l,$$

for any $a \in \mathbb{R}^d$; here, $N_1 \sim \mathcal{N}(0, 1)$. So, since

$$[\Sigma^{-1}(X - nc)]_j^{2l} \leq 2^{2l-1} \left\{ \left(\frac{d}{4\{\lambda_{\min}(\Sigma)\}^2} \right)^l + [\Sigma^{-1}(t - nc)]_j^{2l} \right\}$$

for $t \in [X]$, it follows that

$$\begin{aligned} \mathbb{E}\{[\Sigma^{-1}(W - nc)]_j^{2l}\} &= \sum_{X \in \mathbb{Z}^d} [\Sigma^{-1}(X - nc)]_j^{2l} \int_{[X]} \varphi_n(t) dt \\ &\leq 2^{2l-1} \left\{ \left(\frac{d}{4\{\lambda_{\min}(\Sigma)\}^2} \right)^l + n^l \mathbb{E}\{ \{(e^{(j)})^T \Sigma^{-1/2} N_d\}^{2l} \} \right\} \\ &\leq 2^{2l-1} \left\{ \left(\frac{d}{4\{\lambda_{\min}(\Sigma)\}^2} \right)^l + n^l k(l)(\Sigma^{-1})_{jj}^l \right\}, \end{aligned}$$

and the stated bound follows. Part (b) is similar, but simpler. \square

7.4 Proof of Lemma 5.2

We note first that, from Lemma 5.1(a),

$$\mathbb{E}|W - nc|_\Sigma^i \leq C(i)(nd)^{i/2}, \quad (7.43)$$

if $n \geq 1/\lambda_{\min}(\Sigma)$. For (a), bounding the difference between $\mathbb{E}\{\Delta f(W)^T b I_n^\delta(W)\}$ and $n^{-1} \mathbb{E}\{(f(W)(W - nc)^T \Sigma^{-1} b I_n^\delta(W))\}$, we begin by observing that

$$\begin{aligned} \mathbb{E}\{\Delta_j f(W) I_n^\delta(W)\} &= \sum_{X \in \mathbb{Z}^d} f(X) \{ \mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X] \} I_n^\delta(X) \\ &\quad + \sum_{X \in \mathbb{Z}^d} f(X) \mathbb{P}[W = X - e^{(j)}] \{ I_n^\delta(X - e^{(j)}) - I_n^\delta(X) \}. \end{aligned} \quad (7.44)$$

Because, from the definition of $I_n^\delta(X)$, $|I_n^\delta(X - e^{(j)}) - I_n^\delta(X)| = 1$ requires $|X - e^{(j)} - nc|_\Sigma > n\delta/3$ and $|X - nc|_\Sigma \leq n\delta/3$, or vice versa, the last term in (7.44) is in modulus at most

$$\mathbb{P}[|W - nc|_\Sigma > n\delta/3 - 1/\sqrt{\lambda_{\min}(\Sigma)}] \max_{|X - nc|_\Sigma \leq n\delta/3 + 1/\sqrt{\lambda_{\min}(\Sigma)}} |f(X)|.$$

Thus it follows from (7.43) and a fourth moment Markov inequality that, if $n \geq \max\{1/\lambda_{\min}(\Sigma), 6/(\delta\sqrt{\lambda_{\min}(\Sigma)})\} = \max\{1/\lambda_{\min}(\Sigma), \psi_{\Sigma}(\delta)\}$, then

$$\begin{aligned} & \left| \sum_{X \in \mathbb{Z}^d} f(X) \mathbb{P}[W = X - e^{(j)}] \{I_n^{\delta}(X - e^{(j)}) - I_n^{\delta}(X)\} \right| \\ & \leq \|f\|_{n\delta/2, \infty}^{\Sigma} \mathbb{P}[|W - nc|_{\Sigma} > n\delta/6] \\ & \leq (6/\delta)^4 d^2 C(4) n^{-2} \|f\|_{n\delta/2, \infty}^{\Sigma} \leq d^2 C_1(\delta) n^{-2} \|f\|_{n\delta/2, \infty}^{\Sigma}, \end{aligned} \quad (7.45)$$

where $C_1(\delta) = (6/\delta)^4 C(4) \in \mathcal{K}_{\Sigma}(\delta)$.

For the remainder of (7.44), we write

$$\mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X] = \int_{[X]} \varphi_n(t) D_j(t) dt,$$

where

$$D_j(t) := \exp \left\{ -\frac{1}{2n} \{ -2[\Sigma^{-1}(t - nc)]_j + (\Sigma^{-1})_{jj} \} \right\} - 1.$$

Since $|e^x - 1 - x| \leq \frac{1}{2}x^2 e^{|x|}$, it follows that, for $|X - nc|_{\Sigma} \leq n\delta/3$,

$$\begin{aligned} & \left| D_j(t) - \frac{1}{n} [\Sigma^{-1}(t - nc)]_j \right| \\ & \leq \frac{1}{2n} |(\Sigma^{-1})_{jj}| + \frac{1}{n^2} \{ ([\Sigma^{-1}(t - nc)]_j)^2 + \frac{1}{4} (\Sigma^{-1})_{jj}^2 \} e^{\xi_j(\delta)}, \end{aligned}$$

where

$$\begin{aligned} \xi_j(\delta) &:= \frac{1}{n} \max_{|X - nc|_{\Sigma} \leq n\delta/3} \left\{ |[\Sigma^{-1}(X - nc)]_j| + |(\Sigma^{-1})_{jj}| + \frac{1}{2} d^{1/2} \|\Sigma^{-1}\| \right\} \\ &\leq \frac{1}{3} \|S^{-1/2}\| \delta + \frac{3}{2\lambda_{\min}(\Sigma)} =: \xi^*(\delta), \end{aligned} \quad (7.46)$$

if $n \geq d^{1/2}/\lambda_{\min}(\Sigma)$, true in turn if $n \geq n_1 := (\lambda_{\min}(\Sigma))^{-8/7}$, because $n \geq d^4$. Note also that $n_1 \geq 1/\lambda_{\min}(\Sigma)$. Hence, fixing δ , for such X and for $t \in [X]$,

$$\left| D_j(t) - \frac{1}{n} [\Sigma^{-1}(X - nc)]_j \right| \leq C_2(\delta) n^{-1} (d^{1/2} + n^{-1} [\Sigma^{-1}(X - nc)]_j^2), \quad (7.47)$$

for $C_2(\delta) := 2e^{\xi^*(\delta)}/\lambda_{\min}(\Sigma) \in \mathcal{K}_{\Sigma}(\delta)$, again if $n \geq n_1$. This in turn implies that, for $|X - nc|_{\Sigma} \leq n\delta/3$,

$$\begin{aligned} & \left| \{ \mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X] \} - n^{-1} \mathbb{P}[W = X] [\Sigma^{-1}(X - nc)]_j \right| \\ & \leq C_2(\delta) n^{-1} (d^{1/2} + n^{-1} [\Sigma^{-1}(X - nc)]_j^2) \mathbb{P}[W = X], \end{aligned} \quad (7.48)$$

and hence that

$$\begin{aligned}
& \left| \sum_{X \in \mathbb{Z}^d} f(X) \{ \mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X] \} I_n^\delta(X) \right. \\
& \quad \left. - n^{-1} \mathbb{E} \{ f(W) [\Sigma^{-1}(W - nc)]_j I_n^\delta(W) \} \right| \\
& \leq C_2(\delta) n^{-1} \mathbb{E} \{ d^{1/2} + n^{-1} [\Sigma^{-1}(W - nc)]_j^2 \} \|f\|_{n\delta/2, \infty}^\Sigma. \tag{7.49}
\end{aligned}$$

Now, writing $b = \sum_{j=1}^d b_j e^{(j)}$ and using linearity and Lemma 5.1(c), requiring $n \geq d/\{4(\lambda_{\min}(\Sigma))^2\}$, the inequality (a) follows, if $n \geq \max\{n_{5.2}, \psi_\Sigma(\delta)\}$, where

$$n_{5.2} := \max \{ d^4, n_1, \{4(\lambda_{\min}(\Sigma))^2\}^{-4/3} \}, \tag{7.50}$$

with

$$C_{5.2}^{(1)}(\delta) := C_1(\delta) + C_2(\delta) \{1 + C'(1)(1 + 1/\lambda_{\min}(\Sigma))\}. \tag{7.51}$$

For (b), bounding the difference between $\mathbb{E}\{\Delta f(W)^T B(W - nc) I_n^\delta(W)\}$ and $\mathbb{E}\{f(W) [n^{-1}(W - nc)^T \Sigma^{-1} B(W - nc) - \text{Tr } B] I_n^\delta(W)\}$, we argue in similar style. For $i \neq j$, writing $E^{(ji)} := e^{(j)}(e^{(i)})^T$, we have

$$\begin{aligned}
& \mathbb{E}\{\Delta f(W)^T E^{(ji)}(W - nc) I_n^\delta(W)\} = \mathbb{E}\{\Delta_j f(W)(W_i - nc_i) I_n^\delta(W)\} \\
& = \sum_{X \in \mathbb{Z}^d} f(X)(X_i - nc_i) \{ \mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X] \} I_n^\delta(X) \tag{7.52} \\
& \quad + \sum_{X \in \mathbb{Z}^d} f(X)(X_i - nc_i) \mathbb{P}[W = X - e^{(j)}] \{ I_n^\delta(X - e^{(j)}) - I_n^\delta(X) \}.
\end{aligned}$$

For $n \geq \max\{n_{5.2}, \psi_\Sigma(\delta)\}$, we bound the second element in (7.52) much as for (7.45), using a Markov inequality, Cauchy–Schwarz and Lemma 5.1(a,b), giving

$$\begin{aligned}
& \left| \sum_{X \in \mathbb{Z}^d} f(X)(X_i - nc_i) \mathbb{P}[W = X - e^{(j)}] \{ I_n^\delta(X - e^{(j)}) - I_n^\delta(X) \} \right| \\
& \leq \mathbb{E}\{|W_i - nc_i| I[|W - nc|_\Sigma > n\delta/6]\} \|f\|_{n\delta/2, \infty}^\Sigma \\
& \leq (6/n\delta)^3 \mathbb{E}\{|W_i - nc_i| |W - nc|_\Sigma^3\} \|f\|_{n\delta/2, \infty}^\Sigma \\
& \leq (6/n\delta)^3 \sqrt{\mathbb{E}|W_i - nc_i|^2} \sqrt{\mathbb{E}|W - nc|_\Sigma^6} \|f\|_{n\delta/2, \infty}^\Sigma \\
& \leq (6/\delta)^3 n^{-1} d^{3/2} \sqrt{2(1 + \Sigma_{ii})C(6)} \|f\|_{n\delta/2, \infty}^\Sigma \\
& \leq d^{3/2} C_3(\delta) n^{-1} \|f\|_{n\delta/2, \infty}^\Sigma, \tag{7.53}
\end{aligned}$$

where $C_3(\delta) = (6/\delta)^3 \sqrt{2(1 + \lambda_{\max}(\Sigma))C(6)} \in \mathcal{K}_\Sigma(\delta)$. The first element in (7.52) is treated using (7.48), Cauchy–Schwarz and Lemma 5.1(b,c), giving

$$\begin{aligned}
& \left| \sum_{X \in \mathbb{Z}^d} f(X)(X_i - nc_i) \{ \mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X] \} I_n^\delta(X) \right. \\
& \quad \left. - n^{-1} \mathbb{E} \{ f(W)(W_i - nc_i) [\Sigma^{-1}(W - nc)]_j I_n^\delta(W) \} \right| \\
& \leq C_2(\delta) n^{-1} \mathbb{E} \{ |W_i - nc_i| (d^{1/2} + n^{-1} [\Sigma^{-1}(W - nc)]_j^2) \} \|f\|_{n\delta/2, \infty}^\Sigma \\
& \leq C_2(\delta) n^{-1/2} \sqrt{2(1 + \Sigma_{ii})} \left(d^{1/2} + \sqrt{C'(2)(1 + (\Sigma^{-1})_{ii}^2)} \right) \|f\|_{n\delta/2, \infty}^\Sigma \\
& \leq d^{1/2} C_4(\delta) n^{-1/2} \|f\|_{n\delta/2, \infty}^\Sigma, \tag{7.54}
\end{aligned}$$

with

$$C_4(\delta) := C_2(\delta) \sqrt{2(1 + \lambda_{\max}(\Sigma))} (1 + \sqrt{C'(2)(1 + \lambda_{\min}(\Sigma)^{-2})}) \in \mathcal{K}_\Sigma(\delta).$$

Note that

$$(W_i - nc_i) [\Sigma^{-1}(W - nc)]_j = (W - nc)^T \Sigma^{-1} E^{(ji)} (W - nc).$$

For $i = j$, there is an extra term:

$$\begin{aligned}
& \mathbb{E} \{ \Delta f(W)^T E^{(ii)} (W - nc) I_n^\delta(W) \} \\
& = \sum_{X \in \mathbb{Z}^d} f(X)(X_i - nc_i) \{ \mathbb{P}[W = X - e^{(i)}] - \mathbb{P}[W = X] \} I_n^\delta(X) \\
& \quad + \sum_{X \in \mathbb{Z}^d} f(X)(X_i - nc_i) \mathbb{P}[W = X - e^{(i)}] \{ I_n^\delta(X - e^{(i)}) - I_n^\delta(X) \} \\
& \quad - \sum_{X \in \mathbb{Z}^d} f(X) \mathbb{P}[W = X - e^{(i)}] I_n^\delta(X - e^{(i)}). \tag{7.55}
\end{aligned}$$

Now

$$\begin{aligned}
& \sum_{X \in \mathbb{Z}^d} f(X) \mathbb{P}[W = X - e^{(i)}] I_n^\delta(X - e^{(i)}) \\
& = \mathbb{E} \{ \Delta_i f(W) I_n^\delta(W) \} + \mathbb{E} \{ f(W) I_n^\delta(W) \},
\end{aligned}$$

and $|\mathbb{E} \{ \Delta_i f(W) I_n^\delta(W) \}| \leq \|\Delta f\|_{n\delta/2, \infty}^\Sigma$, giving

$$\begin{aligned}
& \left| \mathbb{E} \{ \Delta f(W)^T E^{(ii)} (W - nc) I_n^\delta(W) \} \right. \\
& \quad \left. - n^{-1} \mathbb{E} \{ f(W)(W - nc)^T \Sigma^{-1} E^{(ii)} (W - nc) I_n^\delta(W) \} - \mathbb{E} \{ f(W) I_n^\delta(W) \} \right| \\
& \leq d^{3/2} C_3(\delta) n^{-1} \|f\|_{n\delta/2, \infty}^\Sigma + d^{1/2} C_4(\delta) n^{-1/2} \|f\|_{n\delta/2, \infty}^\Sigma + \|\Delta f\|_{n\delta/2, \infty}^\Sigma. \tag{7.56}
\end{aligned}$$

The second estimate now follows for general $B = \sum_{i=1}^d \sum_{j=1}^d B_{ij} E^{(ij)}$, by linearity, with

$$C_{5.2}^{(2)}(\delta) := C_4(\delta) + C_3(\delta), \quad (7.57)$$

provided that $n \geq d^2$.

The proof of the final part of Lemma 5.2, bounding the difference between $\mathbb{E}\{\Delta f(W)^T B(W - nc) I_n^\delta(W)\}$ and $\mathbb{E}\{f(W) [n^{-1}(W - nc)^T \Sigma^{-1} B(W - nc) - \text{Tr } B] I_n^\delta(W)\}$, proceeds in very much the same way, but starting with $e^{(j)} b^T$ in place of $E^{(ji)}$ in (7.52) and (7.55), for any $b \in \mathbb{R}^d$. The quantities $(X_i - nc_i)$ and $(W_i - nc_i)$ are replaced in the computations by $b^T(X - nc)$ and $b^T(W - nc) = b^T \Sigma^{1/2} \Sigma^{-1/2}(W - nc)$ respectively. The error terms corresponding to (7.53) and (7.54) then yield the bounds $d^2 C_3'(\delta) |b| n^{-1} \|f\|_{n\delta/2, \infty}^\Sigma$ and $d C_4'(\delta) |b| n^{-1/2} \|f\|_{n\delta/2, \infty}^\Sigma$, with

$$\begin{aligned} C_3'(\delta) &:= (6/\delta)^3 C(4) \sqrt{\lambda_{\max}(\Sigma)}; \\ C_4'(\delta) &:= C_2(\delta) \sqrt{C(2) \lambda_{\max}(\Sigma)} \{1 + \sqrt{C'(2)(1 + \lambda_{\min}(\Sigma)^{-2})}\}. \end{aligned}$$

The analogue of (7.55) yields an error bounded by $|b_j| \|\Delta f\|_{n\delta/2, \infty}^\Sigma$. Part (c) now follows by writing $B = \sum_{j=1}^d e^{(j)} (e^{(j)})^T B$, and applying the above bounds with $b = (e^{(j)})^T B$ for each $1 \leq j \leq d$. \square

7.5 Theorem 1.1 is implied by Theorem 5.5

For Theorem 1.1, we assume that

$$\mathbb{E}|W - nc|^2 \leq dVn; \quad d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \leq \varepsilon_1 \text{ for each } 1 \leq j \leq d, \quad (7.58)$$

and that, for some $\eta \leq \eta_0$ and for all $h: \mathbb{Z}^d \rightarrow \mathbb{R}$,

$$\begin{aligned} &|\mathbb{E}\{\mathcal{A}_n h(W)\} I[|W - nc| \leq n\eta/6]| \\ &\leq \bar{\Lambda}(\varepsilon_{20} \|h\|_{n\eta_0/4, \infty} + \varepsilon_{21} n^{1/2} \|\Delta h\|_{n\eta_0/4, \infty} + \varepsilon_{22} n \|\Delta^2 h\|_{n\eta_0/4, \infty}). \end{aligned} \quad (7.59)$$

Conditions (a) and (b) of Theorem 5.5 are thus clearly satisfied, with $V/\lambda_{\min}(\Sigma)$ for V . For Condition (c), for any $\eta > 0$, let $\eta^- := \eta/\sqrt{\lambda_{\max}(\Sigma)}$, $\eta^+ := \eta/\sqrt{\lambda_{\min}(\Sigma)}$. Then

$$\begin{aligned} \{X \in \mathbb{Z}^d: |X - nc|_\Sigma \leq n\eta^-\} &\subset \{X \in \mathbb{Z}^d: |X - nc| \leq n\eta\} \\ &\subset \{X \in \mathbb{Z}^d: |X - nc|_\Sigma \leq n\eta^+\}. \end{aligned} \quad (7.60)$$

Thus, for any $f: \mathbb{Z}^d \rightarrow \mathbb{R}$ and any $\eta > 0$, we have

$$\|f\|_{n\eta^-, \infty}^\Sigma \leq \|f\|_{n\eta/4, \infty} \leq \|f\|_{n\eta^+, \infty}^\Sigma,$$

where $\|f\|_{n\eta,\infty} := \max_{|X-nc| \leq n\eta} |f(X)|$ and $\|f\|_{n\eta,\infty}^\Sigma$ is as in (4.8). Noting that, for η_0 defined in Theorem 1.1, we have $\eta_0^+ = \tilde{\delta}_0$ for $\tilde{\delta}_0$ as defined in Theorem 5.5, the right hand side of (7.59) is bounded above by

$$\overline{\Lambda} \left(\varepsilon_{20} \|h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma + \varepsilon_{21} n^{1/2} \|\Delta h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma + \varepsilon_{22} n \|\Delta^2 h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma \right). \quad (7.61)$$

On the other hand,

$$\begin{aligned} & | \mathbb{E} \{ \mathcal{A}_n h(W) \} I[|W - nc| \leq n\eta/6] - \mathbb{E} \{ \mathcal{A}_n h(W) \} I[|W - nc|_\Sigma \leq n\eta^-/6] | \\ & \leq | \mathbb{E} \{ \mathcal{A}_n h(W) (I[|W - nc| \leq n\eta/6] - I[|W - nc|_\Sigma \leq n\eta^-/6]) \} | \\ & \leq \frac{1}{2} n \mathbb{E} \{ | \text{Tr}(\sigma^2 \Delta^2 h(W)) | I[n\eta^-/6 < |W - nc|_\Sigma \leq n\eta^+/6] \} \\ & \quad + \mathbb{E} \{ | \Delta h(W)^T A(W - nc) | I[n\eta^-/6 < |W - nc|_\Sigma \leq n\eta^+/6] \} \\ & \leq \frac{1}{2} n \|\Delta^2 h\|_{n\eta^+/6,\infty}^\Sigma \|\sigma^2\|_1 \mathbb{P}[|W - nc|_\Sigma > n\eta^-/6] \\ & \quad + \|\Delta h\|_{n\eta^+/6,\infty}^\Sigma \|A\| \mathbb{E} \{ |W - nc| I[|W - nc|_\Sigma > n\eta^-/6] \}. \end{aligned}$$

Then, if $\mathbb{E}|W - nc|^2 \leq dVn$, we have

$$\mathbb{P}[|W - nc|_\Sigma > n\eta^-/6] \leq \frac{36\mathbb{E}|W - nc|_\Sigma^2}{(n\eta^-)^2} \leq \frac{36\rho(\Sigma)dV}{n\eta^2},$$

and

$$\begin{aligned} & \mathbb{E} \{ |W - nc| I[|W - nc|_\Sigma > n\eta^-/6] \} \\ & \leq \sqrt{\lambda_{\max}(\Sigma)} \mathbb{E} \{ |W - nc|_\Sigma I[|W - nc|_\Sigma > n\eta^-/6] \} \leq 6\rho(\Sigma)dV/\eta, \end{aligned}$$

from the definition of η^- . Hence it follows that

$$\begin{aligned} & | \mathbb{E} \{ \mathcal{A}_n h(W) \} I[|W - nc|_\Sigma \leq n\eta^-/6] | \\ & \leq \frac{18\rho(\Sigma)dV}{n\eta^2} \|\sigma^2\|_1 n \|\Delta^2 h\|_{n\eta^+/6,\infty}^\Sigma + \frac{6\rho(\Sigma)dV}{n^{1/2}\eta} \|A\| n^{1/2} \|\Delta h\|_{n\eta^+/6,\infty}^\Sigma \\ & \quad + \overline{\Lambda} (\varepsilon_{20} \|h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma + \varepsilon_{21} n^{1/2} \|\Delta h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma + \varepsilon_{22} n \|\Delta^2 h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma) \\ & \leq \overline{\Lambda} (\varepsilon_{20} \|h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma + \varepsilon'_{21}(\eta, n) n^{1/2} \|\Delta h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma + \varepsilon'_{22}(\eta, n) n \|\Delta^2 h\|_{n\tilde{\delta}_0/4,\infty}^\Sigma), \end{aligned}$$

with

$$\begin{aligned} \varepsilon'_{21}(\eta, n) &= \varepsilon_{21} + 6n^{-1/2} d(\rho(\Sigma)V/\eta) \|A\|/\overline{\Lambda}; \\ \varepsilon'_{22}(\eta, n) &= \varepsilon_{22} + 18n^{-1} d^{5/2}(V/\eta^2) \|\sigma^2\|/\overline{\Lambda}, \end{aligned}$$

using $\|\sigma^2\|_1 \leq d^{3/2} \|\sigma^2\|$. With the choice $\delta' = \frac{1}{2}\eta^- \leq \frac{1}{2}\eta_0^+ = \frac{1}{2}\tilde{\delta}_0$, it follows that Condition (c) of Theorem 5.5 is satisfied. Thus, from Theorem 5.5, the

conditions of Theorem 1.1 imply the conclusion

$$\begin{aligned}
d_{\text{TV}}(\mathcal{L}(W), \mathcal{DN}_d(nc, n\Sigma)) &\leq C_{5.5}(V/\lambda_{\min}(\Sigma), \eta) \log n \\
&\quad (d^3 n^{-1/2} + d^{7/2}(\bar{\gamma}(\sigma^2)/\bar{\Lambda})\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon'_{21}(\eta, n) + d^{1/2}\varepsilon'_{22}(\eta, n)) \\
&\leq C'(d^3 n^{-1/2} + d^{7/2}(\bar{\gamma}(\sigma^2)/\bar{\Lambda})\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22}) \log n,
\end{aligned}$$

with

$$C' := C_{5.5}(V/\lambda_{\min}(\Sigma), \eta) + (6V/\bar{\Lambda}) \max\{\rho(\Sigma)\eta^{-1}\|A\|, 3\eta^{-2}\|\sigma^2\|\}. \quad \square$$

7.6 Proof of Lemma 6.5

To bound the moments of $Z := (ndv)^{-1/2}\Sigma^{-1/2}(W - \mu)$, we use the equation $\mathbb{E}h(W') - \mathbb{E}h(W) = 0$ for suitably chosen real functions h . First, we take $h(w) = (w - \mu)^T \Sigma^{-1}(w - \mu)$, giving

$$\mathbb{E}\{2\xi^T \Sigma^{-1}(W - \mu) + \xi^T \Sigma^{-1} \xi\} = 0.$$

Noting that $\xi^T \Sigma^{-1} \xi = \text{Tr}(\Sigma^{-1/2} \xi \xi^T \Sigma^{-1/2})$, and using (6.3), we have

$$\begin{aligned}
&-\mathbb{E}\{2n^{-1}(W - \mu)^T \Sigma^{-1} A(W - \mu) + 2n^{-1/2}\|A\|^{1/2} R_1(W)^T \Sigma^{-1}(W - \mu)\} \\
&= \mathbb{E}\text{Tr}(\sigma_{\Sigma}^2(W)) = \text{Tr}(\sigma_{\Sigma}^2),
\end{aligned}$$

where $\sigma_{\Sigma}^2(W) := \Sigma^{-1/2} \sigma^2(W) \Sigma^{-1/2}$. Writing $s_n^2 := \mathbb{E}|Z|^2$, it follows from (6.5) and because $A\Sigma + \Sigma A^T + \sigma^2 = 0$ that

$$2\alpha_1 dv s_n^2 \leq (dv\alpha_1)^{1/2} (\text{Tr}(\sigma_{\Sigma}^2))^{1/2} (1 + s_n^2) + \text{Tr}(\sigma_{\Sigma}^2).$$

From the definition of v , it thus follows directly that $s_n^2 \leq 2$, establishing the first part.

For the third moment, we start with $h(z) = (1 + z^T z)^{3/2}$. The function h has derivatives

$$Dh(z) = 3(1 + z^T z)^{1/2} z$$

and

$$D^2 h(z) = \frac{3zz^T}{(1 + z^T z)^{1/2}} + 3(1 + z^T z)^{1/2} I.$$

Furthermore,

$$\begin{aligned}
&\left| \{h(z + \zeta) - h(z)\} - 3(1 + z^T z)^{1/2} \zeta^T z - \frac{3(\zeta^T z)^2}{2(1 + z^T z)^{1/2}} - \frac{3}{2}(1 + z^T z)^{1/2} |\zeta|^2 \right| \\
&=: d_3(h, z, \zeta) \leq k_{3,h} |\zeta|^3,
\end{aligned} \tag{7.62}$$

for a constant $k_{3,h} \leq 22$ that does not depend on d . This can be seen by considering separately the cases where $|\zeta| \geq (|z| \vee 1)$, $|\zeta| \leq |z|$ and $1 \geq |\zeta| \geq |z|$.

For $|\zeta| \geq (|z| \vee 1)$, simply take the terms one by one, giving

$$d_3(h, z, \zeta) \leq |\zeta|^3(\{5^{3/2} + 2^{3/2}\} + 3 \cdot 2^{1/2} + \frac{3}{2} + \frac{3}{2} \cdot 2^{1/2}) \leq 22|\zeta|^3.$$

For $1 \geq |\zeta| \geq |z|$, use the bounds

$$|(1+x)^{1/2} - 1| \leq \frac{1}{2}x^{1/2}; \quad |(1+x)^{3/2} - 1 - \frac{3}{2}x| \leq \frac{3}{8}x^{3/2}$$

in $0 \leq x \leq 1$ to give

$$|(1+z^T z)^{1/2} - 1| \leq \frac{1}{2}|\zeta|; \quad |h(z+\zeta) - h(z) - \frac{3}{2}(2\zeta^T z + \zeta^T \zeta)| \leq 27|\zeta|^3/8.$$

Then the first, second and fourth terms in $d_3(h, z, \zeta)$ together give at most

$$|\zeta|^3(\frac{27}{8} + \frac{3}{2} + \frac{3}{4}) \leq \frac{45}{8}|\zeta|^3,$$

and the third adds at most $\frac{3}{2}|\zeta|^3$ to this. For $|\zeta| \leq |z|$, Taylor's expansion gives

$$\left| (1+x+y)^{3/2} - (1+x)^{3/2} - \frac{3}{2}y(1+x)^{1/2} - \frac{3y^2}{8\sqrt{1+x}} \right| \leq \frac{|y|^3}{16(1+x)^{3/2}}.$$

We take $x = z^T z$ and $y = 2\zeta^T z + \zeta^T \zeta$, for which $|y| \leq 3|\zeta||z|$. The first, second and fourth terms in $d_3(h, z, \zeta)$ together thus give

$$\frac{3(2\zeta^T z + \zeta^T \zeta)^2}{8(1+z^T z)^{1/2}},$$

up to an error of at most

$$\frac{|2\zeta^T z + \zeta^T \zeta|^3}{16(1+z^T z)^{3/2}} \leq \frac{27|\zeta|^3|z|^3}{16|z|^3} \leq \frac{27}{16}|\zeta|^3.$$

Then

$$\left| \frac{3(2\zeta^T z + \zeta^T \zeta)^2}{8(1+z^T z)^{1/2}} - \frac{3(\zeta^T z)^2}{2(1+z^T z)^{1/2}} \right| \leq \frac{12|\zeta|^3|z| + 3|\zeta|^4}{8|z|} \leq \frac{15}{8}|\zeta|^3,$$

giving an overall bound of $\frac{57}{16}|\zeta|^3$.

We now substitute $z = Z = z(W)$ and $\zeta = (ndv)^{-1/2}\Sigma^{-1/2}\xi$ into (7.62), and take expectations. Since

$$\mathbb{E}h(Z(W + \xi)) = \mathbb{E}h(Z(W)),$$

this immediately gives

$$\begin{aligned}
& \mathbb{E}\{-3(1 + Z^T Z)^{1/2}(ndv)^{-1/2}\xi^T \Sigma^{-1/2} Z\} \\
& \leq n^{-1} \mathbb{E}\left\{\frac{3Z^T \Sigma^{-1/2} \xi \xi^T \Sigma^{-1/2} Z}{2dv(1 + Z^T Z)^{1/2}} + \frac{3}{2}(1 + Z^T Z)^{1/2} \frac{|\Sigma^{-1/2} \xi|^2}{dv}\right\} + \frac{k_{3,h} \chi_\Sigma}{(ndv)^{3/2}} \\
& \leq n^{-1} \frac{3}{2dv} \mathbb{E}\{(2|Z| + 1)|\Sigma^{-1/2} \xi|^2\} + \frac{k_{3,h} \chi_\Sigma}{(ndv)^{3/2}}. \tag{7.63}
\end{aligned}$$

Now

$$\begin{aligned}
& \mathbb{E}\{-3(1 + Z^T Z)^{1/2}(ndv)^{-1/2}\xi^T \Sigma^{-1/2} Z\} \\
& = \mathbb{E}\{-3(1 + Z^T Z)^{1/2}(ndv)^{-1/2}(n^{-1}(W - \mu)^T A^T + n^{-1/2}\|A\|^{1/2} R_1(W)^T) \Sigma^{-1/2} Z\} \\
& = n^{-1} \mathbb{E}\{3(1 + Z^T Z)^{1/2}(\frac{1}{2} Z^T \sigma_\Sigma^2 Z - (dv)^{-1/2}\|A\|^{1/2} R_1(W)^T \Sigma^{-1/2} Z)\}, \tag{7.64}
\end{aligned}$$

and, using (6.6),

$$(dv)^{-1/2}\|A\|^{1/2} \mathbb{E}\{(1 + Z^T Z)^{1/2} |R_1(W)^T \Sigma^{-1/2} Z|\} \leq \frac{1}{4} \alpha_1 (1 + \mathbb{E}|Z|^3). \tag{7.65}$$

Then, by the AM-GM inequality, for any $a > 0$,

$$|Z| |\Sigma^{-1/2} \xi|^2 \leq \frac{1}{3} \{(a|Z|)^3 + 2(a^{-1/2} |\Sigma^{-1/2} \xi|)^3\},$$

so that, taking $a = (dv\alpha_1)^{1/3}$,

$$(dv)^{-1} \mathbb{E}\{|Z| |\Sigma^{-1/2} \xi|^2\} \leq \frac{\alpha_1}{3} \{\mathbb{E}|Z|^3 + 2(n/\|A\|)^{1/2} L_\Sigma\}. \tag{7.66}$$

Combining (7.63)–(7.66), recalling that $\text{Tr}(\sigma_\Sigma^2) = dv\alpha_1$, and multiplying by n , it follows that

$$\begin{aligned}
3\alpha_1 \mathbb{E}|Z|^3 & \leq \frac{3}{4} \alpha_1 (1 + \mathbb{E}|Z|^3) + \alpha_1 \{\mathbb{E}|Z|^3 + 2(n/\|A\|)^{1/2} L_\Sigma\} \\
& \quad + \frac{3}{2} \alpha_1 + k_{3,h} \alpha_1 L_\Sigma \sqrt{\frac{\alpha_1}{\|A\|}},
\end{aligned}$$

giving $\mathbb{E}|Z|^3 \leq 2(1 + 10(n/\|A\|)^{1/2} L_\Sigma)$ if $n/\alpha_1 \geq 1$, because $k_{3,h} \leq 22$. The final inequality is then immediate. \square

7.7 Proofs of Lemma 6.1 and Proposition 6.2

To prove Lemma 6.1, we need to show that, if $X_n^\delta \sim \Pi_n^\delta$ for some $\delta \leq \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$, then $\mathbb{E}\{|X_n^\delta - nc|_\Sigma^2\} = O(n)$. Now, by Dynkin's formula, we have $\mathbb{E}\{\mathcal{A}_n^\delta h(X_n^\delta)\} = 0$ for any choice of h . Take $h(X) = h_0(X) = |X - nc|_\Sigma^2$ as in Lemma 2.2, for which $\mathcal{A}_n^\delta h_0(X) \leq -\alpha_1 h_0(X)$ in $|X - nc|_\Sigma \geq K_{2.2} \sqrt{nd}$.

Then, from (2.8), for all $|X - nc|_\Sigma \leq n \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, $\mathcal{A}_n^\delta h_0(X) \leq ndK'_1$, where, from (2.10), and (2.12),

$$K'_1 := L_0 d^{-1} \text{Tr}(\sigma_\Sigma^2) \leq L_0 \lambda_{\max}(\sigma_\Sigma^2) \leq L_0 \rho(\sigma^2) \rho(\Sigma).$$

This implies that

$$0 = \mathbb{E}\{\mathcal{A}_n^\delta h_0(X_n^\delta)\} \leq -\alpha_1 \mathbb{E}\{|X_n^\delta - nc|_\Sigma^2 I[|X - nc|_\Sigma \geq K_{2.2} \sqrt{nd}]\} + ndK'_1.$$

Since also, using (2.11),

$$\mathbb{E}\{|X_n^\delta - nc|_\Sigma^2 I[|X - nc|_\Sigma < K_{2.2} \sqrt{nd}]\} \leq ndK_{2.2}^2 \leq 4ndL_0 \rho(\sigma^2) \rho(\Sigma),$$

the claim is proved. \square

The proof of Proposition 6.2 is much more involved. We start with a concentration bound, used in the proof of the main estimate, Lemma 6.3, to handle the truncation.

Lemma 7.1. *Define $K_\Sigma := 2\bar{\Lambda}K_{2.5}/(d\theta_1\alpha_1) \in \mathcal{K}$. Under Assumptions G0–G4, for any $0 < \delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, $n \geq n_{2.2}$ and $\eta > K_{2.2}\sqrt{d/n}$,*

$$\Pi_n^\delta\{|X - nc|_\Sigma > n\eta\} \leq \eta^{-2} d^2 K_\Sigma e^{-n\theta_1\eta^2}.$$

In particular, for any fixed $\eta > 0$, $\Pi_n^\delta\{|X - nc|_\Sigma > n\eta\} = O(n^{-r})$ as $n \rightarrow \infty$, for any $r \geq 1$.

Proof. Again, for $\delta \leq \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$ and $X_n^\delta \sim \Pi_n^\delta$, we have $\mathbb{E}\{\mathcal{A}_n^\delta h(X_n^\delta)\} = 0$ for any choice of h , by Dynkin's formula. Take $h(X) = h_{\theta_1}(X)$ as in Lemma 2.2, for which, from (2.17) and for $n \geq n_{2.2}$,

$$\mathcal{A}_n^\delta h_{\theta_1}(X) \leq -\frac{1}{2}n^{-1}\alpha_1\theta_1 h_0(X)h_{\theta_1}(X) \quad \text{for } |X - nc|_\Sigma \geq K_{2.2}\sqrt{nd}.$$

This, with (2.20), implies that

$$\frac{1}{2}n^{-1}\alpha_1\theta_1 \mathbb{E}\{h_0(X)h_{\theta_1}(X)I[|X - nc|_\Sigma \geq K_{2.2}\sqrt{nd}]\} \leq n\Lambda K_{2.5}.$$

Hence, for $\eta > K_{2.2}\sqrt{d/n}$, it follows that

$$\frac{1}{2}n^{-1}\alpha_1\theta_1 (n\eta)^2 e^{n^{-1}\theta_1(n\eta)^2} \Pi_n^\delta\{|X - nc|_\Sigma > n\eta\} \leq n\Lambda K_{2.5},$$

proving the first part. The second is then immediate. \square

The proof of the next lemma is rather close to that of Theorem 3.1, so we only give a quick sketch.

Lemma 7.2. *Under Assumptions G0–G3, for any fixed $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, and for any $J \in \mathcal{J}$,*

$$d_{TV}\{\Pi_n^\delta, \Pi_n^\delta * \varepsilon_J\} = O(n^{-1/2})$$

as $n \rightarrow \infty$. where ε_J denotes the point mass at J and $*$ denotes convolution.

Remark 7.3. *Note that we cannot directly replace J by $e^{(j)}$ here, to obtain Proposition 6.2, because $e^{(j)}$ may not belong to \mathcal{J} . Under Assumption G4, we can do so: see below.*

Proof. Fix any $U > 0$, and use the stationarity of Π_n^δ to give the inequality

$$d_{TV}\{\Pi_n^\delta, \Pi_n^\delta * \varepsilon_J\} \leq \sum_{X \in \mathbb{Z}^d} \Pi_n^\delta(X) d_{TV}\{\mathcal{L}_X(X_n^\delta(U)), \mathcal{L}_X(X_n^\delta(U) + J)\}, \quad (7.67)$$

where \mathcal{L}_X denotes distribution conditional on $\{X_n^\delta(0) = X\}$. By Lemma 6.1, it then follows that, for any $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$,

$$d_{TV}\{\Pi_n^\delta, \Pi_n^\delta * \varepsilon_J\} \leq D_{Jn}(\delta/2) + \frac{4dV_\Sigma}{n\delta^2} = D_{Jn}(\delta/2) + O(n^{-1}), \quad (7.68)$$

where

$$D_{Jn}(\delta') := \sum_{X: |X - nc|_\Sigma \leq n\delta'} \Pi_n^\delta(X) d_{TV}\{\mathcal{L}_X(X_n^\delta(U)), \mathcal{L}_X(X_n^\delta(U) + J)\}.$$

This alters our problem to one of finding a bound of similar form, but now involving the transition probabilities of the process X_n^δ over a finite time U , started in any state X which is reasonably close to nc .

By Assumption G3, J -jumps occur in X_n^δ with rate at least $n\mu_0^J$, whenever it is in the set $\mathcal{X}_n^\delta(J)$. Thus, by analogy with (3.4), we can realize the chain X_n^δ with $X_n^\delta(0) = X_0$ in the form $X_n^\delta(u) := X_0 + JN_n^\delta(u) + W_n^\delta(u)$, where the transition $(l, W) \rightarrow (l+1, W)$ occurs at rate $n\mu_0^J$. This leads to a decomposition

$$\begin{aligned} & d_{TV}\{\mathcal{L}_X(X_n^\delta(U)), \mathcal{L}_X(X_n^\delta(U) + J)\} \\ & \leq \frac{1}{2} \sum_{l \geq 0} |\mathbb{P}_{X_0}[N_n^\delta(U) = l] - \mathbb{P}_{X_0}[N_n^\delta(U) = l-1]| \\ & \quad + \frac{1}{2} \sum_{X \in \mathbb{Z}^d} \sum_{l \geq 1} \mathbb{P}_{X_0}[N_n^\delta(U) = l-1] |q_{l, X_0}^U(X - lJ) - q_{l-1, X_0}^U(X - lJ)|, \end{aligned} \quad (7.69)$$

where $q_{l, X}^U(W)$ is as defined in (3.6).

Much as for (3.9), and using Lemma 2.5, the first sum in (7.69) is bounded by

$$\mathbb{P}_{X_0}[\hat{\tau}_n^\delta \leq U] + \{n\mu_0^J U\}^{-1/2} = O(n^{-1/2}), \quad (7.70)$$

if we choose $U = 1/\sqrt{\Lambda\mu_0^J}$, where $\hat{\tau}_n^\delta$ is as defined in (3.8). For the second part of (7.69), we argue as in the proof of Theorem 3.1, using the Radon–Nikodym derivative $R_{(\mathbf{s}_{l-1}, s_*)}^U(u, w^u) := d\mathbb{P}_{(\mathbf{s}_{l-1}, s_*)}^U / d\mathbb{P}_{\mathbf{s}_{l-1}, X}^U(w^u)$. As long as $R_{(\mathbf{s}_{l-1}, s_*)}^U(u, w^u) \leq 2$ and $u \leq \hat{\tau}_\delta$, the $\mathbb{P}_{\mathbf{s}_{l-1}, X}^U$ -martingale $R_{(\mathbf{s}_{l-1}, s_*)}^U(u, w^u)$ makes jumps of size at most $2|J|L_1/(n\varepsilon_0)$, and this enables the quadratic variation of the stopped martingale to be controlled by $n^{-1}|J|^2 K(\varepsilon_0)\Lambda u$, as in (3.21). Choosing $U = 1/\sqrt{\Lambda\mu_0^J}$ once more, and arguing as for (3.23) and (3.25), the second part in (7.69) is also shown to be of order $O(n^{-1/2})$. Combining these observations with (7.68), the lemma follows. \square

To deduce Proposition 6.2 from Lemma 7.2, take any $1 \leq j \leq d$. Then it is immediate from the triangle inequality that, because $\sum_{l=1}^{r(j)} J_l^{(j)} = e^{(j)}$ for $J_1^{(j)}, \dots, J_{r(j)}^{(j)}$ as given in Assumption G4, we also have

$$d_{\text{TV}}(\Pi_n^\delta, \Pi_n^\delta * \varepsilon_{e^{(j)}}) \leq \sum_{l=1}^{r(j)} d_{\text{TV}}(\Pi_n^\delta, \Pi_n^\delta * \varepsilon_{J_l^{(j)}}) = O(n^{-1/2}). \quad \square$$

References

- [1] A. D. BARBOUR (1988) Stein’s method and Poisson process convergence. *J. Appl. Probab.* **25(A)**, 175–184.
- [2] A. D. BARBOUR, L. HOLST & S. JANSON (1992) *Poisson Approximation*. Oxford Univ. Press.
- [3] A. D. BARBOUR & P. K. POLLETT (2012) Total variation approximation for quasi-equilibrium distributions, II. *Stoch. Procs. Applics* **122**, 3740–3756.
- [4] A. D. BARBOUR & A. XIA (1999) Poisson perturbations. *ESAIM: P&S* **3**, 131–150.
- [5] V. BENTKUS (2003) On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113**, 385–402.
- [6] L. H. Y. CHEN (1975) Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

- [7] L. H. Y. CHEN, L. GOLDSTEIN & Q.-M. SHAO (2011) *Normal approximation by Stein's method*. Springer–Verlag, Berlin.
- [8] X. FANG (2014) Discretized normal approximation by Stein's method. *Bernoulli* **20**, 1404–1431.
- [9] X. FANG & A. RÖLLIN (2014) Rates of convergence for multivariate normal approximation with application to dense graphs and doubly indexed permutation statistics. *Bernoulli* (to appear).
- [10] J. G. KEMENY & J. L. SNELL (1960) *Finite Markov chains*. van Nostrand, Princeton NJ.
- [11] J. G. KEMENY & J. L. SNELL (1961) Finite continuous time Markov chains. *Teor. Verojatnost. i Primenen.* **6**, 101–105.
- [12] H. K. KHALIL (2002) *Nonlinear systems*. 3rd Edn, Prentice Hall, New Jersey.
- [13] T. LINDVALL (2002) *Lectures on the coupling method*. 2nd Edn, Dover Publications, Mineola NY.
- [14] E. L. PRESMAN (1983) On the approximation of binomial distributions by means of infinitely divisible ones. *Teor. Verojatnost. i Primenen.* **28**, 393–403.
- [15] G. REINERT & A. RÖLLIN (2009) Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37**, 2150–2173.
- [16] Y. RINOTT & V. ROTAR (1996) A multivariate CLT for local dependence with $n^{-1/2} \log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.* **56**, 333–350.
- [17] A. RÖLLIN (2005) Approximation of sums of conditionally independent random variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.
- [18] A. RÖLLIN (2007) Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17**, 1596–1614.
- [19] A. RÖLLIN & N. ROSS (2015) Local limit theorems via Landau–Kolmogorov inequalities. *Bernoulli* **21**, 851–880.

- [20] G. O. ROBERTS & J. S. ROSENTHAL (1996) Quantitative bounds for convergence rates of continuous time Markov processes. *Elec. J. Probab.* **1**, Paper no. 9, 21pp.
- [21] S. N. SOCOLL & A. D. BARBOUR (2010) Translated Poisson approximation to equilibrium distributions of Markov population processes. *Meth. Comp. Appl. Probab.* **12**, 567–586.
- [22] C. STEIN (1986) *Approximate computation of expectations*. IMS Lecture Notes **7**, Hayward, California.